

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 08:57:39 ; Search time 236.64 seconds  
(without alignments)  
-82.217 Million cell updates/sec

Title: US-08-945-805-1  
Perfect score: 20  
Sequence: 1 cctgaaggatttcctcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 285916

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vi:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vi:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	14.2	71.0	21	5	AR052910 Sequence
2	14.2	71.0	21	5	AR054273 Sequence
3	14.2	71.0	21	5	AR054475 Sequence
4	14.2	71.0	36	5	AR066266 Sequence
c 5	13.2	66.0	27	5	155896 Sequence 25
6	13.2	66.0	36	5	AR066265 Sequence
7	12.8	64.0	20	5	AR067063 Sequence
8	12.8	64.0	29	5	AR053312 Sequence
9	12.8	64.0	29	5	140931 Sequence 46
c 10	12.6	63.0	26	7	S71804S1
11	12.6	63.0	33	5	AR066260 Sequence
c 12	12.6	63.0	33	5	AR066260 Sequence
13	12.6	63.0	43	5	AR055645 Sequence
14	12.4	62.0	26	5	AR00959 Sequence
15	12.4	62.0	26	5	164727 Sequence 7
c 16	12.2	61.0	20	5	A42353 Sequence 13
c 17	12.2	61.0	20	5	A44386 Sequence 16
c 18	12.2	61.0	20	5	A47173 Sequence 16
c 19	12.2	61.0	20	5	A56647 Sequence 14
c 20	12.2	61.0	20	5	A80368 Sequence 14
c 21	12.2	61.0	20	5	184726 Sequence 14
c 22	12.2	61.0	39	5	A67638 Sequence 58
c 23	12	60.0	26	5	A82429 Sequence 17
c 24	12	60.0	36	14	SYNMBPB1
c 25	12	60.0	36	14	SYNMBPB2
26	11.8	59.0	21	5	119927 Sequence 24
c 27	11.8	59.0	28	5	A46164 Sequence 2
c 28	11.8	59.0	49	5	AR053248 Sequence
29	11.8	59.0	50	16	SV4EV14B
30	11.6	58.0	30	5	AR000137 Sequence
31	11.6	58.0	30	5	166251 Sequence 3
c 32	11.6	58.0	48	10	HSU14097
33	11.6	58.0	49	5	E13038
c 34	11.6	58.0	50	9	HUMSAU3A13
c 35	11.4	57.0	24	5	A49177 Sequence 11
c 36	11.2	56.0	20	5	AR029477 Sequence
37	11.2	56.0	20	5	190250 Sequence 8
38	11.2	56.0	23	5	A82762 Sequence 50
c 39	11.2	56.0	33	5	AR020756 Sequence
c 40	11.2	56.0	33	5	134037 Sequence 11
41	11.2	56.0	35	10	HSPCRPMR
42	11.2	56.0	36	5	X65342 H.sapiens d
c 43	11.2	56.0	36	5	AR066269 Sequence
c 44	11.2	56.0	47	5	168145 Sequence 18
c 45	11.2	56.0	47	5	AR032505 Sequence
					129245 Sequence 11

ALIGNMENTS

```
RESULT 1
AR052910      21 bp  DNA
DEFINITION    Sequence 34 from patent US 5833976.
ACCESSION     AR052910
VERSION       AR052910.1  GI:5977772
KEYWORDS      .
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 21)
AUTHORS      Malefyt,R.,deWaal, Howard,M., Hsu,D., Ishida,H., O'Garra,A.,
              Spits,H. and Zlotnik,A.
TITLE         Use of interleukin-10 (IL-10) to treat endotoxin- or
              superantigen-induced toxicity
JOURNAL       Patent: US 5833976-A 34 10-NOV-1998;
FEATURES      Location/Qualifiers
              source
              1..21
              /organism="unknown"
BASE COUNT    3 a 9 c 6 g 3 t
ORIGIN

Query Match   71.0%; Score 14.2; DB 5; Length 21;
Best Local Similarity 84.2%; Pred. No. 3.3e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctc 19
    ||| ||||| |||||
Db 3 CCTGGAAGGATCTCCCC 21

RESULT 2
AR054273      21 bp  DNA
DEFINITION    Sequence 34 from patent US 5837232.
ACCESSION     AR054273
VERSION       AR054273.1  GI:5979850
KEYWORDS      .
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 21)
AUTHORS      De Waal Malefyt,R., Howard,M., Hsu,D., Ishida,H., O'Garra,A.,
              Spits,H. and Zlotnik,A.
TITLE         Use of an interleukin-10 antagonist to treat a B cell mediated
              autoimmune disorder
JOURNAL       Patent: US 5837232-A 34 17-NOV-1998;
FEATURES      Location/Qualifiers
              source
              1..21
              /organism="unknown"
BASE COUNT    3 a 9 c 6 g 3 t
ORIGIN

Query Match   71.0%; Score 14.2; DB 5; Length 21;
Best Local Similarity 84.2%; Pred. No. 3.3e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctc 19
    ||| ||||| |||||
Db 3 CCTGGAAGGATCTCCCC 21

RESULT 3
AR054475      21 bp  DNA
DEFINITION    Sequence 34 from patent US 5837293.
ACCESSION     AR054475
VERSION       AR054475.1  GI:5980052
KEYWORDS      .
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 21)
AUTHORS      Hook,M., Jonsson,K., Patti,J.M. and Gurusiddappa,S.
              MHC II analog from Staphylococcus aureus
TITLE         Patent: US 5648240-A 25 15-JUL-1997;
JOURNAL
FEATURES      Location/Qualifiers
              source
              1..27
              /organism="unknown"
BASE COUNT    7 a 8 c 7 g 5 t
ORIGIN

ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 21)
AUTHORS      De Waal Malefyt,R., Howard,M., Hsu,D., Ishida,H., O'Garra,A.,
              Spits,H. and Zlotnik,A.
TITLE         Use of interleukin-10 analogs for antagonists to treat endotoxin-
              or superantigen-induced toxicity
JOURNAL       Patent: US 5837293-A 34 17-NOV-1998;
FEATURES      Location/Qualifiers
              source
              1..21
              /organism="unknown"
BASE COUNT    3 a 9 c 6 g 3 t
ORIGIN

Query Match   71.0%; Score 14.2; DB 5; Length 21;
Best Local Similarity 84.2%; Pred. No. 3.3e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctc 19
    ||| ||||| |||||
Db 3 CCTGGAAGGATCTCCCC 21

RESULT 4
AR066266      36 bp  DNA
DEFINITION    Sequence 31 from patent US 5849900.
ACCESSION     AR066266
VERSION       AR066266.1  GI:5996482
KEYWORDS      .
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 36)
AUTHORS      Moelling,K.
TITLE         Inhibition of viruses by antisense oligomers capable of binding to
              polypurine rich tract of single-stranded RNA or RNA-DNA hybrids
JOURNAL       Patent: US 5849900-A 31 15-DEC-1998;
FEATURES      Location/Qualifiers
              source
              1..36
              /organism="unknown"
BASE COUNT    4 a 12 c 12 g 8 t
ORIGIN

Query Match   71.0%; Score 14.2; DB 5; Length 36;
Best Local Similarity 84.2%; Pred. No. 3.3e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cttgaaggattccctcc 20
    || ||||| ||||| ||
Db 8 CTGGAAGGGTTTTCCTC 26

RESULT 5
I55896/c      27 bp  DNA
DEFINITION    Sequence 25 from patent US 5648240.
ACCESSION     I55896
VERSION       I55896.1  GI:2476690
KEYWORDS      .
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 27)
AUTHORS      Hook,M., Jonsson,K., Patti,J.M. and Gurusiddappa,S.
              MHC II analog from Staphylococcus aureus
TITLE         Patent: US 5648240-A 25 15-JUL-1997;
JOURNAL
FEATURES      Location/Qualifiers
              source
              1..27
              /organism="unknown"
BASE COUNT    7 a 8 c 7 g 5 t
ORIGIN
```

## ORIGIN

Query Match 66.0%; Score 13.2; DB 5; Length 27;  
Best Local Similarity 83.3%; Pred. No. 1.1e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cttgaaggattccctc 19  
||||| ||||| |||

Db 26 CTGAAGTGATTGCCGC 9

## RESULT 6

AR066265 LOCUS 36 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 30 from patent US 5849900.  
ACCESSION AR066265  
VERSION AR066265.1 GI:5996481  
KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 36)  
AUTHORS Moelling,K.

TITLE Inhibition of viruses by antisense oligomers capable of binding to polypurine rich tract of single-stranded RNA or RNA-DNA hybrids

JOURNAL Patent: US 5849900-A 30 15-DEC-1998;

FEATURES Location/Qualifiers

source 1..36

/organism="unknown"

BASE COUNT 3 a 11 c 12 g 10 t

## ORIGIN

Query Match 66.0%; Score 13.2; DB 5; Length 36;  
Best Local Similarity 83.3%; Pred. No. 1.1e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttgaaggattccctcc 20  
||||| ||||| |||

Db 9 TGAAGGCTTTCCCTTC 26

## RESULT 7

AR067063 LOCUS 20 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 411 from patent US 5851760.  
ACCESSION AR067063  
VERSION AR067063.1 GI:5998285  
KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 20)

AUTHORS Evans,G.A. and Smith,M.W.

TITLE Method for generation of sequence sampled maps of complex genomes

JOURNAL Patent: US 5851760-A 411 22-DEC-1998;

FEATURES Location/Qualifiers

source 1..20

/organism="unknown"

BASE COUNT 4 a 5 c 5 g 6 t

## ORIGIN

Query Match 64.0%; Score 12.8; DB 5; Length 20;  
Best Local Similarity 87.5%; Pred. No. 1.9e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tgaaggattccctc 19  
||| || ||||| |||

Db 2 TGAGGCGATTCCCTC 17

## RESULT 8

AR053312 LOCUS 29 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 46 from patent US 5834209.  
ACCESSION AR053312  
VERSION AR053312.1 GI:5978174  
KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 29)

AUTHORS Korsmeyer,S.J.

TITLE Bcl-x/bcl-2 associated cell death regulator

JOURNAL Patent: US 5834209-A 46 10-NOV-1998;

FEATURES Location/Qualifiers

source 1..29

/organism="unknown"

BASE COUNT 4 a 7 c 8 g 10 t

## ORIGIN

Query Match 64.0%; Score 12.8; DB 5; Length 29;  
Best Local Similarity 87.5%; Pred. No. 1.9e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcc 16  
|||| | |||| | |||

Db 12 CCTTCAAGGGACTTCC 27

## RESULT 9

I40931 LOCUS 29 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 46 from patent US 5622852.  
ACCESSION I40931  
VERSION I40931.1 GI:2082411  
KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 29)

AUTHORS Korsmeyer,S.J.

TITLE Bcl-x/Bcl-2 associated cell death regulator

JOURNAL Patent: US 5622852-A 46 22-APR-1997;

FEATURES Location/Qualifiers

source 1..29

/organism="unknown"

BASE COUNT 4 a 7 c 8 g 10 t

## ORIGIN

Query Match 64.0%; Score 12.8; DB 5; Length 29;  
Best Local Similarity 87.5%; Pred. No. 1.9e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcc 16  
|||| | |||| | |||

Db 12 CCTTCAAGGGACTTCC 27

## RESULT 10

S71804S1/c LOCUS 26 bp DNA PLN 07-MAY-1993  
DEFINITION {autonomous Activator element (AC), clone CC32} [Lycopodium  
esculentum-tomatoes, cv. VF36, transgenic RO A78, transposon,  
Transposon, 26 nt, segment 1 of 2].

ACCESSION S71804

VERSION S71804.1 GI:240749

KEYWORDS

SOURCE 1 of 2

ORGANISM tomato cv. VF36 transposon transgenic RO A78.

Lycopodium esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;



Query Match 62.0%; Score 12.4; DB 5; Length 26;  
Best Local Similarity 92.9%; Pred. NO. 3.1e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 tgaagggtttccc 17  
| | | | | | | | | |  
Db 8 TGAGGGTTTCCC 21

RESULT 15  
I64727  
LOCUS I64727 26 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 7 from patent US 5665591.  
ACCESSION I64727  
VERSION I64727.1 GI:2481621  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 26)  
AUTHORS Sonenshein,G.E., Lawrence,R. and Bellas,R.E.  
TITLE Regulation of smooth muscle cell proliferation  
JOURNAL Patent: US 5665591-A 7 09-SEP-1997;  
FEATURES  
source  
1..26  
/organism="unknown"  
BASE COUNT 7 a 8 c 5 g 6 t  
ORIGIN

Query Match 62.0%; Score 12.4; DB 5; Length 26;  
Best Local Similarity 92.9%; Pred. NO. 3.1e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 tgaagggtttccc 17  
| | | | | | | | | |  
Db 8 TGAGGGTTTCCC 21

Search completed: May 23, 2000, 10:36:29  
Job time: 5930 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:41:18 ; Search time 96.59 Seconds  
(without alignments)  
51.805 Million cell updates/sec

Title: us-08-945-805-1

Perfect score: 20

Sequence: 1 ccttgaagggttcctccc 20

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 391736

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	1 T69675	Transcription fact
2	20	100.0	36	1 Q90101	VCAM-1 expression
3	20	100.0	47	1 Q90100	VCAM-1 expression
4	18	90.0	19	1 Q90103	VCAM-1 expression
5	15	75.0	34	1 Q90105	VCAM-1 expression
6	14.2	71.0	21	1 Q37156	Probe to detect G-
7	14.2	71.0	21	1 V08007	Probe G-CSF for In
8	14.2	71.0	30	1 T45759	Human granulocyte-
9	14.2	71.0	36	1 Q45406	Oligonucleotide fo
10	14	70.0	34	1 Q90105	VCAM-1 expression
11	13.2	66.0	27	1 T84443	Staphylococcal MHC
12	13.2	66.0	36	1 Q45405	Oligonucleotide fo
13	12.8	64.0	20	1 Q82411	Chromosome 11 (loc
14	12.8	64.0	29	1 T29469	Bad cDNA fragment,
15	12.8	64.0	32	1 T64663	G-CSF receptor ago
16	12.8	64.0	36	1 T64640	G-CSF receptor ago
17	12.8	64.0	36	1 V55465	Granulocyte-colony
18	12.6	63.0	22	1 T44297	5'- and 3'-Guanosi
19	12.6	63.0	22	1 X24151	c-fos directed pho
20	12.6	63.0	30	1 Q21824	Reverse transcript
21	12.6	63.0	33	1 Q45400	Oligonucleotide fo
22	12.6	63.0	33	1 Q45400	Oligonucleotide fo
23	12.4	62.0	26	1 T89626	Mutant nuclear fac
24	12.4	62.0	26	1 T29699	C-myc URE mutant N
25	12.2	61.0	20	1 Q88728	Human c-fos modifi
26	12.2	61.0	20	1 T44440	Antisense oligonuc
27	12.2	61.0	20	1 T44243	c-fos antisense co
28	12.2	61.0	20	1 X24197	Phosphonomoester
29	12.2	61.0	20	1 X33914	c-fos expression i
30	12.2	61.0	21	1 T44296	3'-Guanosine-cappe
31	12.2	61.0	21	1 T42150	c-fos directed pho
32	12.2	61.0	25	1 Q87379	PCR primer 9f (MOG
33	12.2	61.0	31	1 T79309	Esterase MLITC-29L
34	12.2	61.0	34	1 T30852	Human TGF-beta typ

#### ALIGNMENTS

RESULT 1

T69675

ID T69675 standard; DNA; 20 BP.

AC T69675;

DT 04-AUG-1997 (first entry)

DE Transcription factor NF-kappa-B DNA binding site antagonist.

KW Decoy; antagonist; NF-kappa-B; NF-KB; transcription; regulation;

KW prevention; treatment; disease; ischaemia; ischaemia; inflammation;

KW autoimmune; cancer; metastasis; cachexia; organ; transplantation;

KW surgery; ds.

OS Synthetic.

PN W09635430-A1.

PD 14-NOV-1996.

PF 10-MAY-1996; J01234.

PR 12-MAY-1995; JP-114990.

PR 02-NOV-1995; JP-285504.

PA (FUJI ) FUJISAWA PHARM CO LTD.

PI Chiba T, Kawamura I, Maeda K, Morishita R, Ogiwara T;

PI Sugimoto T;

DR WPI; 96-518400/51.

PT Anti-sense NF-kB agent - for treatment of ischaemia, inflammatory

PT disease auto-immune disease, etc.

PS Claim 7; Page 9; 18pp; Japanese.

CC The present sequence is a decoy, which specifically antagonises

CC the nucleic acid site to which a NF-kappa-B transcription regulator

CC binds. It can be used to prevent or treat diseases caused by

CC NF-kappa-B, e.g. ischaemia, inflammatory and autoimmune disease,

CC cancer metastasis and cachexia, especially following organ

CC transplant or surgery.

SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T;

Query Match 100.0%; Score 20; DB 1; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.095;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttgaagggttcctccc 20

|||||

Db 1 CcTtGaAGGgATtTCCcTCC 20

RESULT 2

Q90101

ID Q90101 standard; DNA; 36 BP.

AC Q90101;

DT 11-JAN-1996 (first entry)

DE VCAM-1 expression inhibiting oligonucleotide.

KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;

KW transcriptional regulatory factor; diagnosis; treatment; restenosis;

KW atherosclerosis; inflammatory disease; ds.

OS Synthetic.

PN W09512415-A1.

PD 11-MAY-1995.

PF 07-NOV-1994; U12797.

PR 05-NOV-1993; US-147878.

PA (ISIS-) ISIS PHARM INC.

PA (UYEM-) UNIV EMORY.

PCR primer used to  
Human granulocyte  
Primer for human T  
PCR primer used fo  
Porcine KIT exon/1  
Reverse primer KIT  
Fibroblast growth  
potato citrate syn  
platelet aggregati  
phage lambda cro g  
32P-labelled oligo

c 35 12.2 61.0 39 1 V16058  
c 36 12.2 61.0 41 1 X19303  
c 37 12.2 61.0 47 1 X36362  
c 38 12 60.0 24 1 X08409  
c 39 12 60.0 26 1 V80707  
c 40 12 60.0 26 1 X32896  
c 41 12 60.0 32 1 V21392  
c 42 11.8 59.0 17 1 V96448  
c 43 11.8 59.0 20 1 Q50354  
c 44 11.8 59.0 21 1 Q57447  
c 45 11.8 59.0 21 1 V80095

```

PI Bennett CF, Medford RM;
DR WPI; 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17; Page 33; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 36 BP; 3 A; 13 C; 9 G; 11 T;

Query Match 100.0%; Score 20; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctcc 20
    |||||
Db 16 CCTTGAAGGATTCCCTCC 35

RESULT 3
Q90100
ID Q90100 standard; DNA; 47 BP.
AC Q90100;
DT 11-JAN-1996 (first entry)
DE VCAM-1 expression inhibiting oligonucleotide.
KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;
KW transcriptional regulatory factor; diagnosis; treatment; restenosis;
KW atherosclerosis; inflammatory disease; ds.
OS Synthetic.
PN WO9512415-A1.
PD 11-MAY-1995.
PF 07-NOV-1994; U12797.
PR 05-NOV-1993; US-147878.
PA (ISIS-) ISIS PHARM INC.
PA (UYEM-) UNIV EMORY.
PI Bennett CF, Medford RM;
DR WPI; 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17; Page 33; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 47 BP; 3 A; 17 C; 12 G; 15 T;

Query Match 100.0%; Score 20; DB 1; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctcc 20
    |||||
Db 22 CCTTGAAGGATTCCCTCC 41

RESULT 4
Q90103
ID Q90103 standard; DNA; 19 BP.
AC Q90103;
DT 11-JAN-1996 (first entry)
DE VCAM-1 expression inhibiting oligonucleotide.
KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;
KW transcriptional regulatory factor; diagnosis; treatment; restenosis;
KW atherosclerosis; inflammatory disease; ds.

```

```

OS Synthetic.
PN WO9512415-A1.
PD 11-MAY-1995.
PF 07-NOV-1994; U12797.
PR 05-NOV-1993; US-147878.
PA (ISIS-) ISIS PHARM INC.
PA (UYEM-) UNIV EMORY.
PI Bennett CF, Medford RM;
DR WPI; 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17; Page 34; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 19 BP; 3 A; 5 C; 5 G; 6 T;

Query Match 90.0%; Score 18; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ttgaaggattccctcc 20
    |||||
Db 1 TTGAAGGATTCCCTCC 18

RESULT 5
Q90105
ID Q90105 standard; DNA; 34 BP.
AC Q90105;
DT 11-JAN-1996 (first entry)
DE VCAM-1 expression inhibiting oligonucleotide.
KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;
KW transcriptional regulatory factor; diagnosis; treatment; restenosis;
KW atherosclerosis; inflammatory disease; ds.
OS Synthetic.
PN WO9512415-A1.
PD 11-MAY-1995.
PF 07-NOV-1994; U12797.
PR 05-NOV-1993; US-147878.
PA (ISIS-) ISIS PHARM INC.
PA (UYEM-) UNIV EMORY.
PI Bennett CF, Medford RM;
DR WPI; 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17; Page 35; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 34 BP; 7 A; 9 C; 7 G; 11 T;

Query Match 75.0%; Score 15; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gaaggattccctcc 19
    |||||
Db 1 GAAGGATTCCCTC 15

RESULT 6
Q37156

```

IC Q37156 standard; DNA; 21 BP.  
AC Q37156;  
DT 23-JUN-1993 (first entry)  
DE Probe to detect G-CSF sequences.  
KW Granulocyte Colony Stimulating Factor;  
KW cytokine synthesis inhibitor; inflammation;  
KW monokine production; Southern analysis; ss.  
OS Synthetic.  
PN WO9302693-A.  
PD 18-FEB-1993.  
PF 06-AUG-1992; U06378.  
PR 06-AUG-1991; US-742129.  
PA (SCHE ) SCHERING CORP.  
PI De Waal Malefyt R, Howard M, Hsu DH, Ishida H, Ogarra A;  
PI Spits H, Zlotnik A;  
DR WPI; 93-076172/09.  
PT Use of interleukin-10 to modulate inflammation or T-cell mediated  
PT immune function - for treating septic and toxic shock,  
PT auto-immune diseases, tumours and infectious diseases  
PS Example B6; Page 85; 208pp; English.  
CC Northern and Southern hybridisations were performed to determine  
CC the level at which IL-10 and IL-4 inhibit monokine production. The  
CC probe Q37156 was used in Southern analysis of PCR products to detect  
CC G-CSF coding sequences. The sequence of the probe corresponds  
CC to nucleotides 400-421 of the sequence given in Nature 319; 415  
CC 1986. It was found that IL-1 alpha, IL-6, TNF alpha, GM-CSF and  
CC G-CSF expression was strongly inhibited by IL-10 and IL-4 at the mRNA  
CC level. IL-1 beta and IL-8 expression was only slightly affected by  
CC IL-10.  
SQ Sequence 21 BP; 3 A; 9 C; 6 G; 3 T;

Query Match 71.0%; Score 14.2; DB 1; Length 21;  
Best Local Similarity 84.2%; Pred. No. 95;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccttgaaggatttcctc 19  
||| ||||| |||||  
Db 3 CCTGGAAGGATCTCCCC 21

RESULT 7  
V08007  
ID V08007 standard; DNA; 21 BP.  
AC V08007;  
DT 20-JAN-1999 (first entry)  
DE Probe G-CSF for Interleukin-10 coding sequence.  
KW Interleukin-10; IL-10; septic shock; bacterial infection; toxic shock;  
KW infectious shock; inflammation; immune response modulation; therapy;  
KW probe; ss.  
OS Synthetic.  
PN US5833976-A.  
PD 10-NOV-1998.  
PF 24-MAR-1995; 410654.  
PR 06-AUG-1992; US-926853.  
PR 06-AUG-1991; US-742129.  
PR 19-APR-1994; US-229854.  
PR 24-MAR-1995; US-410654.  
PA (SCHE ) SCHERING CORP.  
PI De Waal Malefyt R, Howard M, Hsu D, Ishida H, Ogarra A,  
PI Spits H, Zlotnik A;  
DR WPI; 99-008644/01.  
PT Treating shock conditions from e.g. bacterial infections - comprises  
PT administering interleukin-10  
PS Example 14; Column 42; 109pp; English.  
CC This sequence represents a probe for an interleukin-10 (IL-10) coding  
CC sequence. The IL-10 protein can be used in the method of the invention  
CC for ameliorating a symptom of: (a) septic shock in a host suffering from  
CC a bacterial (preferably gram negative) infection; (b) toxic shock;  
CC (c) infectious shock; or (d) inflammation. The method comprises  
CC administering a biologically active IL-10 (preferably human) protein,  
CC analogue or a fragment (preferably full length). The treatment is used to  
CC modulate immune responses caused by the different shock syndromes, which

CC are endotoxin or superantigen induced toxicity, or autoimmune related  
CC conditions. The conditions are side-effects of microbial infections,  
CC caused by release of their protein products, especially on anti-microbial  
CC treatment, which when cells are killed, they lyse, releasing proteins  
CC which induce the shock conditions. IL-10 inhibits TNF-alpha (tumour  
CC necrosis factor-alpha) and TNF-gamma synthesis, which as part of an  
CC immune response elicits the shock syndromes.  
SQ Sequence 21 BP; 3 A; 9 C; 6 G; 3 T;

Query Match 71.0%; Score 14.2; DB 1; Length 21;  
Best Local Similarity 84.2%; Pred. No. 95;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccttgaaggatttcctc 19  
||| ||||| |||||  
Db 3 CCTGGAAGGATCTCCCC 21

RESULT 8  
T45759  
ID T45759 standard; DNA; 30 BP.  
AC T45759;  
DT 17-FEB-1997 (first entry)  
DE Human granulocyte-colony stimulating factor gene probe.  
KW Polymerase chain reaction; PCR; interleukin; IL; cytokine;  
KW growth factor; animal model; stem cell; haematopoiesis; CD34;  
KW infection; HIV; human immunodeficiency virus; immunomodulator;  
KW immortalise; bone marrow; stromal cell; engraftment; determination;  
KW study; research; ss.  
OS Homo sapiens.  
PN WO9617627-A2.  
PD 13-JUN-1996.  
PF 08-DEC-1995; U15986.  
PR 09-DEC-1994; US-352957.  
PA (GART/) GARTNER S.  
PA (HALL/) HALL E.  
PA (KAUS/) KAUSHAL S.  
PA (KESS/) KESSLER S.  
PA (LRUS/) LA RUSSA V.  
PA (MOSC/) MOSCA J D.  
PI Gartner S, Hall E, Kaushal S, Kessler S, La Russa V;  
PI Mosca JD;  
DR WPI; 96-286928/29.  
PT Animal models for human haematopoiesis - have en-grafted human or  
PT primate stem cells in the presence of immortalised bone marrow  
PT stromal cells  
PS Example 1; Page 15; 43pp; English.  
CC T45754-T45762 are probes used to determine whether or not the  
CC cytokines IL-1, IL-6, IL-8, GM-CSF, G-CSF, M-CSF, TGF-alpha and  
CC stem cell factor are expressed by a human bone marrow stromal  
CC cell line, Lof(11-10). The cells were found to produce the cytokines  
CC which support the growth of CD34+ stem cells. The Lof(11-10) cells  
CC were injected into SCID mice (previously irradiated to provide an  
CC internal space for CD34+ cells to populate). Five to seven days  
CC after injection the mice were injected with 3 to 5 human CD34+ cells.  
CC After 3 weeks, human CD34+ cells were found in the bone marrow of  
CC the mice. The immortalised bone marrow stem cells create a human  
CC microenvironment supplying human cytokines in the animals to provide  
CC for the engraftment, maintenance and differentiation of CD34+ stem  
CC cells. Animal models created by administering Lof(11-10) cells are  
CC used to study and determine the effectiveness of therapies against  
CC disease such as HIV infection. They can also be used to assay for  
CC haematopoietic growth factors, immunomodulators and/or immune  
CC toxins.  
SQ Sequence 30 BP; 5 A; 11 C; 10 G; 4 T;

Query Match 71.0%; Score 14.2; DB 1; Length 30;  
Best Local Similarity 84.2%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccttgaaggatttcctc 19

Db 9 CTTGGAAGGGATCTCCCC 27  
 ||| ||||| |||| |

## RESULT 9

Q45406

ID Q45406 standard; DNA; 36 BP.

AC Q45406;

DT 11-NOV-1994 (first entry)

DE Oligonucleotide forming triplex with viral polypurine tract.

KW HIV; human immunodeficiency virus; retrovirus; hepatitis virus;

KW reverse transcription; virus replication; inhibition; treatment;

KW therapy; polypurine; triplex; antisense; ss.

OS Synthetic.

PN W09407367-A.

PD 14-APR-1994.

PF 29-SEP-1993; U09300.

PR 29-SEP-1992; US-954184.

PA (APOL-) APOLLON INC.

PA (PLAC-) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PI Moelling K.

DR WPI; 94-135099/16.

PT Antiviral oligomers that bind poly-purine tracts of

PT single-stranded RNA or RNA - DNA hybrids - used to target the

PT early stages of viral replication before double stranded DNA is

PT formed

PS Example 4; Page 30; 54pp; English.

CC Administration of antisense or triplex forming oligonucleotides

CC which bind polypurine tracts (PPT) may be used in the therapy or

CC treatment of individuals infected with retroviruses or hepatitis

CC viruses since in these two families of viruses, two primers are

CC involved in the reverse transcription of viral RNA into double

CC stranded DNA, one of which is a PPT. The antisense or triplex

CC forming oligonucleotides can inhibit the early stages of viral

CC replication by binding to the PPT primer or by binding to PPT

CC tracts in the RNA-DNA hybrid molecule formed after reverse

CC transcription of the viral RNA. This oligonucleotide was incubated

CC with an in vitro transcribed 5' end labelled PKJ2 RNA of 134

CC nucleotides in length to which a 40-mer deoxyribonucleotide

CC complementary to the PPT had been hybridised. The presence of this

CC sequence led to protection of the PKJ2 RNA-DNA hybrid from RNase H

CC digestion which would suggest the formation of a triplex. Triplex

CC formation was confirmed using a primer extension technique. A

CC primer binding downstream of the PPT was synthesised and extended in

CC vitro by reverse transcriptase in the presence of

CC oligodeoxynucleotides including one which was radioactively labelled.

CC The newly synthesised DNA was terminated at the site of the PPT when

CC the triplex was formed and blocked extension. Triplex formation

CC would be expected to interfere with viral replication in vivo.

CC See Q45381-Q45417. This is a variant of the sequence described in

CC Q45387.

SQ Sequence 36 BP; 4 A; 12 C; 12 G; 8 T;

Query Match 71.0%; Score 14.2; DB 1; Length 36;

Best Local Similarity 84.2%; Pred. No. 1e+02;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cttgaaggattccctcc 20

|| ||||| ||||| |

Db 8 CTGGAAGGGTTTCCCTTC 26

## RESULT 10

Q90105/c

ID Q90105 standard; DNA; 34 BP.

AC Q90105;

DT 11-JAN-1996 (first entry)

DE VCAM-1 expression inhibiting oligonucleotide.

KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;

KW transcriptional regulatory factor; diagnosis; treatment; restenosis;

KW atherosclerosis; inflammatory disease; ds.

OS Synthetic.

PN W09512415-A1.

PD 11-MAY-1995.

PF 07-NOV-1994; U12797.

PR 05-NOV-1993; US-147878.

PA (ISIS-) ISIS PHARM INC.

PA (UYEM-) UNIV EMORY.

PI Bennett CF, Medford RM;

DR WPI; 95-193802/25.

PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule

PT expression by binding a transcription regulatory element - used to

PT diagnose and treat atherosclerosis, restenosis or inflammatory

PT disease

PS Claim 17; Page 35; 49pp; English.

CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1

CC gene transcriptional regulatory factor, therefore inhibiting the

CC expression of VCAM-1. They can be used in the diagnosis and

CC treatment of restenosis, atherosclerosis and inflammatory

CC diseases.

SQ Sequence 34 BP; 7 A; 9 C; 7 G; 11 T;

Query Match

Best Local Similarity 70.0%; Score 14; DB 1; Length 34;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gaaggattccctc 18

|| ||||| ||||| |

Db 34 GAAGGATTTCCT 21

## RESULT 11

T84443/C

ID T84443 standard; DNA; 27 BP.

AC T84443;

DT 16-NOV-1997 (first entry)

DE Staphylococcal MHC class II antigen analogue gene primer ZPCR9.

KW MHC II analogue protein; major histocompatibility complex;

KW Staphylococcus aureus; adhesin; virulence factor; vaccine;

KW polymerase chain reaction; PCR; primer; vector; pQE20; ss.

OS Synthetic.

PN US5648240-A.

PD 15-JUL-1997.

PF 24-MAY-1994; 248021.

PR 24-MAY-1994; US-248021.

PA (TEXA) UNIV TEXAS A &amp; M.

PI Gurusiddappa S, Hook M, Jonsson K, Patti JM;

DR WPI; 97-372059/34.

PT DNA encoding Staphylococcus aureus broad spectrum adhesin - for

PT production of recombinant adhesin for use in vaccines

PS Example 7; Column 7; 30pp; English.

CC Primer ZPCR6 (T84442) and primer ZPCR9 (T84443) were used to

CC amplify a 2 kb DNA segment of Staphylococcal aureus FDA 574

CC chromosomal DNA. The primers respectively contain BamHI and SalI

CC cleavage sites. The PCR-amplified DNA fragment was inserted into

CC the BamHI/SalI cleavage sites of plasmid pQE30 to generate vector

CC pQE20. This vector can be used to express recombinant

CC staphylococcal MHC II analogue protein in transformed host (pref.

CC E. coli) cells for use in vaccines against staphylococcal

CC infection.

SQ Sequence 27 BP; 7 A; 8 C; 7 G; 5 T;

Query Match

Best Local Similarity 66.0%; Score 13.2; DB 1; Length 27;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cttgaaggattccctc 19

|| ||||| ||||| |

Db 26 CTGGAAGTATTTCGCC 9

## RESULT 12

Q45405

ID Q45405 standard; DNA; 36 BP.  
AC Q45405; 1994 (first entry)  
DT Oligonucleotide forming triplex with viral polypurine tract.  
DE HIV; human immunodeficiency virus; retrovirus; hepatitis virus;  
KW reverse transcription; virus replication; inhibition; treatment;  
KW therapy; polypurine; triplex; antisense; ss.  
OS Synthetic.  
PN W09407367-A.  
PD 14-APR-1994.  
PF 29-SEP-1993; U093300.  
PR 29-SEP-1993; US-954184.  
PA (APOL-) APOLLON INC.  
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
PI Moelling K;  
DR WPI; 94-135099/16.  
PT Antiviral oligomers that bind polypurine tracts of  
PT single-stranded RNA or RNA - DNA hybrids - used to target the  
PT early stages of viral replication before double stranded DNA is  
PT formed  
PS Example 4; Page 30; 54pp; English.  
CC Administration of antisense or triplex forming oligonucleotides  
CC which bind polypurine tracts (PPT) may be used in the therapy or  
CC treatment of individuals infected with retroviruses or hepatitis  
CC viruses since in these two families of viruses, two primers are  
CC involved in the reverse transcription of viral RNA into double  
CC stranded DNA, one of which is a PPT. The antisense or triplex  
CC forming oligonucleotides can inhibit the early stages of viral  
CC replication by binding to the PPT primer or by binding to PPT  
CC tracts in the RNA-DNA hybrid molecule formed after reverse  
CC transcription of the viral RNA. This oligonucleotide was incubated  
CC with an in vitro transcribed 5' end labelled pKJ2 RNA of 134  
CC nucleotides in length to which a 40-mer deoxyribonucleotide  
CC complementary to the PPT had been hybridised. The presence of this  
CC sequence led to protection of the pKJ2 RNA-DNA hybrid from RNase H  
CC digestion which would suggest the formation of a triplex. A  
CC formation was confirmed using a primer extension technique. A  
CC primer binding downstream of the PPT was synthesised and extended in  
CC vitro by reverse transcriptase in the presence of  
CC oligodeoxynucleotides including one which was radioactively labelled.  
CC The newly synthesised DNA was terminated at the site of the PPT when  
CC the triplex was formed and blocked extension. Triplex formation  
CC would be expected to interfere with viral replication in vivo.  
CC See Q45381-Q45417. This is a variant of the sequence described in  
CC Q45387.  
SQ Sequence 36 BP; 3 A; 11 C; 12 G; 10 T;

Query Match 66.0%; Score 13.2; DB 1; Length 36;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttgaaggatttcctcc 20  
| | | | | | | | | | |  
Db 9 TGAAGGATTTCCCTTC 26

RESULT 13  
Q82411  
ID Q82411 standard; DNA; 20 BP.  
AC Q82411;  
DT 11-SEP-1995 (first entry)  
DE Chromosome 11 (locus D11S1190) STS primer CSRL-5f1-TA.  
KW sequence sampled mapping; genomic analysis; complex genome mapping; ss.  
KW cosmid library; chromosome 11; sequence tagged site; STS analysis; ss.  
OS Synthetic.  
PN W09429486-A.  
PD 22-DEC-1994.  
PF 15-JUN-1994; U06810.  
PR 15-JUN-1994; US-078471.  
PR 07-SEP-1993; US-117952.  
PA (SALK ) SALK INST BIOLOGICAL STUDIES.  
PI Evans GA, Smith MW;

DR WPI; 95-036508/05.  
PT Sequencing complex genomes, present as fragments in a cosmid  
PT library - by sequencing end-specific nucleotides of each clone  
PT then correlating with spatial relationship of cosmid, esp. for  
PT mammalian chromosomes.  
PS Example 4; Page 80; 128pp; English.  
CC Sequences were determined from the ends of chromosome 11-specific  
CC cosmids by automated sequencing without intermediate subcloning.  
CC A sample of 371 DNA sequence fragments were determined and of  
CC these, 277 were suitable for STS primer prediction by computer  
CC analysis (using the "Primer" program available from E.Lander, MIT).  
CC The STSs and cosmids were mapped by in situ hybridisation, somatic  
CC cell hybrid analysis or both. Using this method, 370 STSs specific  
CC for human chromosome 11 were generated and most of them were  
CC regionally mapped. This procedure illustrates a novel method for  
CC sequencing complex genomes, designated "sequence sampled mapping".  
CC The sequence sampled mapping method is useful for the completion of  
CC high density sequence-based maps, and ultimately, for the complete  
CC sequencing of genomic DNA directly from cosmid clones.  
CC See Q82001-Q82706 for STS primers. (Also see Q91325-58).  
SQ Sequence 20 BP; 4 A; 5 C; 5 G; 6 T;

Query Match 64.0%; Score 12.8; DB 1; Length 20;  
Best Local Similarity 87.5%; Pred. No. 5e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tgaaggatttcctcc 19  
| | | | | | | | | | |  
Db 2 TGAGGGATTTCCCTTC 17

RESULT 14  
T29469  
ID T29469 standard; cDNA; 29 BP.  
AC T29469;  
DT 03-JAN-1997 (first entry)  
DE Bad cDNA fragment, bases 471-499.  
KW Epitope; murine; bcl-x(L)/bcl-2 associated death promoter; Bad; stroke;  
KW polypeptide; bcl-x; cell death; regulate; BH1; BH2; apoptotic cell death;  
KW cytokine deprivation; IL-3 dependent cell line; immunodeficiency; AIDS;  
KW neurodegenerative disease; senescence; ischaemia; neoplasia; ss.  
OS Mus musculus.  
PN W09613614-A1.  
PD 09-MAY-1996.  
PF 31-OCT-1995; U14246.  
PR 31-OCT-1994; US-333565.  
PA (UNIW ) UNIV WASHINGTON.  
PI Korsmeyer SJ;  
DR WPI; 96-251465/25.  
PT Polynucleotide encoding bcl-x(L)/bcl-2 associated death promoter -  
PT useful to treat neoplasia and apoptosis and to identify agents  
PT inhibiting its binding to bcl-2 or bcl-x(L) to form heteromultimers  
PS Claim 11; Page 104; 130pp; English.  
CC The sequences given in T29454-77 represent fragments derived from the  
CC murine bcl-x(L)/bcl-2 associated death promoter (Bad) gene (see  
CC also T29479). Bad is a 22.1 kD protein which interacts with bcl-2 and  
CC bcl-x proteins and regulates cell death. It has homology to the bcl-2-  
CC related family clustered in the BH1 and BH2 domain. Bad has been found  
CC to hybridise to bcl-x(L) and bcl-2 in yeast two-hybrid assays and in  
CC vivo in mammalian cells. Overexpressed Bad counters the death inhibitory  
CC activity of bcl-x(L), but is much less effective at countering the death  
CC inhibitory activity of bcl-2. Bad expression can accelerate apoptotic  
CC cell death induced by cytokine deprivation in an IL-3 dependent cell  
CC line expressing bcl-x(L), and its also counters the death repressor  
CC activity of bcl-x(L). Bad competes with Bax for binding to bcl-x(L).  
CC Bad may be used to identify agents which inhibit its binding to bcl-2  
CC or bcl-x(L) to form heterodimers. Such agents may be used to treat  
CC neurodegenerative diseases, immunodeficiency diseases, e.g. AIDS,  
CC senescence or ischaemia.  
SQ Sequence 29 BP; 4 A; 7 C; 8 G; 10 T;

Query Match 64.0%; Score 12.8; DB 1; Length 29;  
Best Local Similarity 87.5%; Pred. No. 5.3e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcc 16  
    |||||  
Db 12 CCTTCAAGGGACTTCC 27

RESULT 15  
T64663  
ID T64663 standard; DNA: 32 BP.  
AC T64663;  
DT 17-JAN-1998 (first entry)  
DE G-CSF receptor agonist gene primer 91start.  
KW Granulocyte colony stimulating factor receptor; agonist; G-CSF;  
KW haematopoietic disorder; neutropenia; bone marrow suppression;  
KW stem cell expansion; gene therapy; circular permutation;  
KW polymerase chain reaction; PCR; primer; ss.  
OS Synthetic.  
PN WO9712977-A1.  
PD 10-APR-1997.  
PF 04-OCT-1996; U15935.  
PR 05-OCT-1995; US-004832.  
PA (SEAR ) SEARLE & CO G D.  
PI Braford-Goldberg SR, Feng Y, Klein BK, McKearn JP;  
PI McWhorter CA, Zurfluh LL;  
DR WPI: 97-244718/22.  
PT Modified human granulocyte colony stimulating factor - useful as  
PT G-CSF receptor agonist for treating haematopoietic disorders, e.g.  
PT neutropenia or bone marrow suppression  
PS Example 34: Page 31: 186pp: English.  
CC This synthetic oligonucleotide comprises primer 91start that was  
CC used in the construction of novel gene pMON3464 (see T64628)  
CC encoding a circularly permuted variant (see W15056) of human  
CC granulocyte colony stimulating factor (G-CSF) that acts as a  
CC G-CSF receptor agonist. Such receptor agonists can be used in  
CC methods for stimulating production of haematopoietic cells, ex  
CC vivo expansion of stem cells, treatment of haematopoietic disorders  
CC and human gene therapy.  
SQ Sequence 32 BP: 7 A; 9 C; 10 G; 6 T;

Query Match 64.0%; Score 12.8; DB 1; Length 32;  
Best Local Similarity 87.5%; Pred. No. 5.3e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcc 16  
    |||||  
Db 17 CCTGGAAGGGATATCC 32

Search completed: May 23, 2000, 11:21:29  
Job time: 6011 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:21:58 ; Search time 1337.41 Seconds  
(without alignments)  
60.613 Million cell updates/sec

Title: US-08-945-805-1  
Perfect score: 20  
Sequence: 1 ccttgaaggattccctcc 20  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues  
Total number of hits satisfying chosen parameters: 23194

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

- EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

- 45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	#	Query
--------	---	-------

No.	Score	Match	Length	DB	ID	Description
1	14.2	71.0	31	50	AI697078	AI697078 wc85g12.x
2	13.2	66.0	29	20	TI7517	TI7517 gsr m46 The
c	3	13	65.0	44	D67701	D67701 CELK076GXF
	4	12.2	61.0	31	N34910	N34910 yy48b09.s1
c	5	12	60.0	40	38	AA779179 zj43c07.s
	6	11.4	57.0	40	79	AW249802 2819841.3
c	7	11.4	57.0	43	23	R94622 yq38d08.s1
	8	11.4	57.0	43	62	AI917027 ts5ia12.x
c	9	11.2	56.0	43	66	AI423979 tf36h03.x
	10	11.2	56.0	46	29	AA152923 mr89h01.x
c	11	11.2	56.0	46	48	AI581293 tal0d04.r
	12	11	55.0	49	23	H16724 ym22f11.s1
c	13	10.8	54.0	31	48	AI587478 tr51c05.x
	14	10.8	54.0	37	35	AA594384 ni93e10.s
c	15	10.8	54.0	37	50	AI683766 tw53g05.x
	16	10.8	54.0	46	40	AA916309 on64a06.s
c	17	10.8	54.0	48	79	AW250438 2822406.3
	18	10.6	53.0	31	48	AI587478 tr51c05.x
c	19	10.6	53.0	37	48	AI558529 fb73c11.y
	20	10.6	53.0	37	63	AI938847 sc61b08.y
c	21	10.6	53.0	40	45	AI377260 te59h12.x
	22	10.6	53.0	40	45	AJ237180 vW12d11.r
c	23	10.6	53.0	46	38	AA789528 sw12d11.r
	24	10.6	53.0	46	48	AI582875 ts07a07.x
c	25	10.6	53.0	50	21	D45786 HUMGS02999
	26	10.4	52.0	31	24	N22606 yw31e07.s1
c	27	10.4	52.0	31	47	AI539033 tp76e02.x
	28	10.4	52.0	37	40	AA996065 os33c11.s
c	29	10.4	52.0	43	39	AA872050 o11ld08.s
	30	10.4	52.0	43	47	AI544860 fb69b01.x
c	31	10.4	52.0	46	36	AA642483 ns26d02.s
	32	10.4	52.0	46	40	AA937239 oJ02c06.s
c	33	10.4	52.0	47	41	AU012006 AU012006
	34	10.4	52.0	49	40	AA910591 ok57c09.s
c	35	10.4	52.0	49	60	AI797849 wh79h03.x
	36	10.2	51.0	34	48	AI568952 tq71h03.x
c	37	10.2	51.0	36	23	R85295 yq24e08.s1
	38	10.2	51.0	40	45	AJ237180 AJ237180
c	39	10.2	51.0	46	40	AA960633 ub60a05.s
	40	10.2	51.0	46	43	AI223998 qx12h07.x
c	41	10.2	51.0	49	27	AA002380 mq43a05.r
	42	10	50.0	31	43	AI198881 qf66c10.x
c	43	10	50.0	31	61	AI813711 wj65h12.x
	44	10	50.0	32	21	T63705 yci6e07.r1
c	45	10	50.0	34	44	AI287726 qul2h11.x

ALIGNMENTS

RESULT	1
AI697078	
LOCUS	AI697078 31 bp mRNA EST 17-DEC-1999
DEFINITION	wc85g12.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2325478 3' similar to TR:075369 075369 BETA-FILAMIN. ;, mRNA sequence.
ACCESSION	AI697078
VERSION	AI697078.1 GI:4984978
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 31)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) On Mar 10, 1998 this sequence version replaced gi:2948790. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov

Tissue procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.	
CDNA Library Preparation: M. Bento Soares, Ph.D.	
CDNA Library Arraying: Greg Lennon, Ph.D.	
DNA Sequencing by: Washington University Genome Sequencing Center	
Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html	
Insert Length: 1093 Std Error: 0.00	
Seq primer: -400P from GIBCO	
High quality sequence stop: 1.	
Location/Qualifiers	
FEATURES	Source
1..31	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:2325478"
	/clone_lib="NCI_CGAP_Co3"
	/sex="pooled"
	/tissue_type="colon"
	/lab_host="DH10B"
/note="Vector: pT7F3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7F3 vector. Library went through one round of normalization."	
BASE COUNT	5 a 9 c 9 g 8 t
ORIGIN	
Query Match 71.0%; Score 14.2; DB 50; Length 31;	
Best Local Similarity 84.2%; Pred. No. 2.1e+03;	
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	1 ccttgaaggaggttcctc 19
Db	7 CCTGAAGGACTGCCAC 25
RESULT	2
TI7517	
LOCUS	TI7517 29 bp DNA EST 06-JUN-1994
DEFINITION	gsr m46 The blue guys library Saccharomyces cerevisiae genomic sequence upstream of LacZ fusion similar to MER1, mRNA sequence.
ACCESSION	TI7517.1 GI:457899
VERSION	EST.
KEYWORDS	baker's yeast.
SOURCE	Saccharomyces cerevisiae
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE	1 (bases 1 to 29)
AUTHORS	Burns,N., Griwade,B., Ross-Macdonald,P.B., Choi,E.-Y., Finberg,K., Roeder,G.S. and Snyder,M.
TITLE	Large-scale analysis of gene expression, protein localization and gene disruption in Saccharomyces cerevisiae
JOURNAL	Genes Dev. 8, 1087-1105 (1994)
MEDLINE	95011603
COMMENT	Contact: Roeder GS Department of Biology Yale University KBR 804, P.O. Box 208103, New Haven, CT 06520-8103 Tel: 2034323501 Fax: 2034326161 Email: shirleen_roeder@quickmail.yale.edu Fusion: codon 32 of MER1. LacZ fusion; sporulation / meiosis induced expression. Disruption phenotype: as published. Beta-gal fusion localization pattern: none. Time of Beta-gal fusion localization: none. Seq primer: LacZ sequences in transposon.

FEATURES  
source

Location/Qualifiers

1. .29  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
/clone\_lib="The blue guys library"  
/lab\_host="E.coli"

/note="vector: PRECmtn; A yeast genomic DNA library was prepared in the vector pHS56, and subjected to transposon mutagenesis with mtn3. This mini-transposon carries lacZ sequences that lack an initiation codon; expression of lacZ is only provided by in frame fusion to yeast coding sequence. The yeast genomic DNA carrying the transposon was excised from pHS56 and transplanted back onto the yeast chromosome. Yeast colonies expressing lacZ were screened for in a color assay. A plasmid containing the genomic DNA/lacZ fusion junction was recovered from each individual yeast colony that expressed lacZ activity. These recovered plasmids comprise 'The blue guys library'. The fusion junction was then sequenced to identify the expressed ORF upstream of the fusion."

BASE COUNT 3 a 6 c 5 g 15 t  
ORIGIN

## Query Match

Best Local Similarity 66.0%; Score 13.2; DB 20; Length 29;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 ttgaaggattccctcc 20

||||| | |||||

Db 9 TTGAATTCCTTCCCTCC 26

## RESULT 3

D67701/c

LOCUS

DEFINITION D67701 44 bp mRNA EST 07-DEC-1995  
Caenorhabditis elegans. clone yk76g10 5', mRNA sequence.

ACCESSION D67701.1 GI:1105489

VERSION

KEYWORDS

SOURCE

ORGANISM

Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 44)

Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and

Tabara, H.

Toward an expression map of the C.elegans genome

Unpublished (1994)

On Apr 14, 1993 this sequence version replaced gi:693680.

Contact: Yuji Kohara

Gene Library Lab

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykoha@lab.nig.ac.jp

Insert length: 415 Std Error: 0.00

High quality sequence stop: 369.

## FEATURES

source

1. .44  
/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/db\_xref="taxon:6239"

/clone="yk76g10"

/clone\_lib="Yuji Kohara unpublished cDNA"

/sex="hermaphrodite, male"

/tissue\_type="whole animal"

/dev\_stage="varied"

15 a 11 c 5 g 13 t

## BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 65.0%; Score 13; DB 24; Length 44;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ttgaaggatttc 15

|||||

Db 21 TTGAAGGATTTC 9

## RESULT 4

LOCUS

DEFINITION N34910 31 bp mRNA EST 16-JAN-1996  
YY4809.s1 Soares\_multiple\_sclerosis\_2NBHSP Homo sapiens CDNA  
clone IMAGE:276761 3', similar to SW\_NUCM\_BOVIN P23709  
NADH-UBIQUINONE OXIDOREDUCTASE 30 KD SUBUNIT PRECURSOR ;, mRNA  
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 31)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

On Apr 14, 1993 this sequence version replaced gi:785784.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence starts: 1

High quality sequence stops: 1

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: ml3 -40 forward

High quality sequence stop: 1.

Location/Qualifiers

1. .31

/organism="Homo sapiens"

/db\_xref="GDB:3895137"

/db\_xref="taxon:9606"

/clone="IMAGE:276761"

/clone\_lib="Soares\_multiple\_sclerosis\_2NBHSP"

/sex="male"

/tissue\_type="multiple sclerosis lesions"

/dev\_stage="Age 46"

/lab\_host="DH10B (ampicillin resistant)"

/note="vector: pT73D (Pharmacia) with a modified

polylinker V.TYPE: phagemid; Site\_1: Not I; Site\_2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo(dT)

primer [5',

TGTTACCAATCTGAAGTGGAGCGCGCGATTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis

lesions from one patient was kindly provided by Dr. Kevin

G. Becker (NINDS/NIH). "

BASE COUNT 8 a 7 c 8 g 7 t 1 others

Matches	15;	Conservative	0;	Mismatches	5;	Indels	0;	Gaps	0;
QY	1	cttgaaggattccctcc	20						
Db	17	CCTAGATGCGATTTCTTCC	36						
RESULT	6								
AW249802/c									
LOCUS									
DEFINITION		AW249802	40 bp	mRNA	EST			07-JAN-2000	
		2819841.3	prime NIH_MGC_7	Homo sapiens	cDNA clone	IMAGE:2819841	3'		
ACCESSION		AW249802							
VERSION		AW249802.1	GI:6592795						
KEYWORDS		EST.							
SOURCE		human.							
ORGANISM		Homo sapiens							
REFERENCE		1	(bases 1 to 40)						
AUTHORS		NIH-MGC	http://www.ncbi.nlm.nih.gov/MGC/.						
TITLE		National Institutes of Health, Mammalian Gene Collection (MGC)							
JOURNAL		Unpublished (1999)							
COMMENT		On May 18, 1998 this sequence version replaced gi:3137758.							
		Other ESTs: 2819841.5	prime						
		Contact: Robert Strausberg, Ph.D.							
		Tel: (301) 496-1550							
		Email: Robert.Strausberg@nih.gov							
		Tissue Procurement: DCTD/DTP	cDNA Library Preparation: Ling						
		Hong/Rubin Laboratory	cDNA Library Arrayed by: The I.M.A.G.E.						
		Consortium (LLNL) DNA Sequencing by: Berkeley MGC Sequencing							
		Project Clone Distribution: MGC clone distribution information can							
		be found through the I.M.A.G.E. Consortium/LLNL at:							
		www-bio.llnl.gov/bbrp/image/image.html	Base Calling / Quality						
		Scores: PHRED from University of Washington Genome Center. Vector							
		Trimming: cross_match from University of Washington Genome Center							
		PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley							
		Drosophila Genome Project. University of Washington Genome Center:							
		http://www.genome.washington.edu	Low Quality Sequence: 15						
		contiguous PHRED high quality bases following vector sequence. Very							
		Low Quality Sequence: trace file contained 40 contiguous distinct							
		peaks following vector sequence. Polyadenylation: Based upon the							
		presence of a XhoI site followed by a run of 14 or more T residues							
		at the beginning of the sequence, this cDNA insert was							
		polyadenylated.							
		Plate: LLCM2	row: K	column: 10					
		High quality sequence stop: 15.							
FEATURES		Location/Qualifiers							
source		1..40							
		/organism="Homo sapiens"							
		/db_xref="taxon:9606"							
		/clone="IMAGE:2819841"							
		/clone_lib="NIH_MGC_7"							
		/tissue_type="small cell carcinoma"							
		/cell_line="MGC3"							
		/lab_host="DH10B (phage-resistant)"							
		/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:							
		ECOR1; cDNA made by oligo-dT priming. Directionally							
		cloned into ECOR1/XhoI sites using the following 5'							
		adaptor: GGCACGAG(G). Size-selected >500bp for average							
		insert size 1.8kb. Library constructed by Ling Hong in							
		the laboratory of Gerald M. Rubin (University of							
		California, Berkeley) using ZAP-cDNA synthesis kit							
		(Stratagene) and Superscript II RT (Life Technologies)."							
BASE COUNT		11 a	6 c	9 g	14 t				
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
QY	8	gggattccctcc	20						

```

Db 32 CCGGTTTCCCTCC 20

RESULT 7
LOCUS R94622 43 bp mRNA EST 30-AUG-1995
DEFINITION Yq38d08.s1 Soares fetal liver spleen lNfLS Homo sapiens cDNA clone IMAGE:198063 3' similar to gb:M68895_rnal ALCOHOL DEHYDROGENASE 6 (HUMAN);, mRNA sequence.
ACCESSION R94622
VERSION R94622.1 GI:970017
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 43)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Sep 21, 1992 this sequence version replaced gi:279347.
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 868
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 868 Std Error: 0.00
Seq primer: Promega -2lm13
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..43
/organism="Homo sapiens"
/db_xref="GDB:376713"
/db_xref="taxon:9606"
/clone="IMAGE:198063"
/clone_lib="Soares fetal liver spleen lNfLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAAATTAATAAGATCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 12 a 6 c 7 g 16 t 2 others
ORIGIN

Query Match 57.0%; Score 11.4; DB 23; Length 43;
Best Local Similarity 85.7%; Pred. No. 5.8e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cttgaagggtttc 15
|||||
Db 24 CTTGANGGGTTTC 37

RESULT 8
LOCUS AI917027/c 43 bp mRNA EST 28-JUL-1999
DEFINITION tsial2.x1 NCI_CGAP_Kid8 Homo sapiens cDNA clone IMAGE:232094 3' similar to TR:008816 008816 N-WASP PROTEIN. ;contains TAR1.b2 MSR1 repetitive element ;, mRNA sequence.
ACCESSION AI917027
VERSION AI917027.1 GI:5636882
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 43)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3138634.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Trace considered overall poor quality
Seq primer: -400P from Gibco
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..43
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:232094"
/clone_lib="NCI_CGAP_Kid8"
/tissue_type="renal cell tumor"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.2 Kb. Life Technologies catalog #: 11524-014"
BASE COUNT 12 a 21 c 8 g 2 t
ORIGIN

Query Match 57.0%; Score 11.4; DB 62; Length 43;
Best Local Similarity 92.3%; Pred. No. 5.8e+04;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 gggatttcctcc 20
|||||
Db 35 GGGTTTCCCTCC 23

RESULT 9
LOCUS AI423979 43 bp mRNA EST 09-MAR-1999
DEFINITION tf36h03.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2098325 3' similar to SW:SP49_HUMAN Q15427 SPLICEOSOME ASSOCIATED PROTEIN 49 ;contains element MER22 repetitive element ;, mRNA sequence.
ACCESSION AI423979
VERSION AI423979.1 GI:4269910
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 43)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3138634.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Trace considered overall poor quality
Seq primer: -400P from Gibco
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..43
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:232094"
/clone_lib="NCI_CGAP_Kid8"
/tissue_type="renal cell tumor"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.2 Kb. Life Technologies catalog #: 11524-014"
BASE COUNT 12 a 21 c 8 g 2 t
ORIGIN

Query Match 57.0%; Score 11.4; DB 62; Length 43;
Best Local Similarity 92.3%; Pred. No. 5.8e+04;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 gggatttcctcc 20
|||||
Db 35 GGGTTTCCCTCC 23

```



```

FEATURES
  source
    Location/Qualifiers
      1. .46
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:2043655"
        /dev_stage="8-9 weeks"
        /lab_host="DH10B"
        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
        polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
        was prepared from mRNA obtained from pooled 8-9 week
        (TGTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3')].
        Double-stranded cDNA was ligated to Eco RI adaptors
        (Pharmacia), digested with Not I and cloned into the Not I
        and Eco RI sites of the modified pT7T3 vector. Library
        went through one round of normalization, and was
        constructed by Bento Soares and M. Fatima Bonaldo. "
      13 a 10 c 19 g 4 t
      BASE COUNT 13 a 10 c 19 g 4 t
      ORIGIN
        Query Match 56.0%; Score 11.2; DB 48; Length 46;
        Best Local Similarity 81.2%; Pred. No. 7.4e+04;
        Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

      Qy 5 gaagggtattccctcc 20
        ||| |||||
      Db 33 GAGGGTTTTCCTCC 18

      RESULT 12
      H16724/c 49 bp mRNA EST 29-JUN-1995
      LOCUS
      DEFINITION Ym22f11.s1 Soares infant brain INIB Homo sapiens cDNA clone
      IMAGE:48850 3' similar to gb:M63175 AUTOCRINE MOTILITY FACTOR
      RECEPTOR PRECURSOR (HUMAN);, mRNA sequence.
      ACCESSION H16724.1 GI:882964
      VERSION
      KEYWORDS EST.
      SOURCE
      ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Primates; Catarrhini; Hominidae; Homo.
      1 (bases 1 to 49)
      REFERENCE
      AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
      Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
      Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
      Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
      Willson,R.
      TITLE The WashU-Merck EST Project
      JOURNAL Unpublished (1995)
      COMMENT On May 8, 1995 this sequence version replaced gi:800068.
      Contact: Wilson RK
      Washington University School of Medicine
      4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
      Tel: 314 286 1800
      Fax: 314 286 1810
      Email: est@watson.wustl.edu
      Insert Size: 1054
      High quality sequence starts: 1
      High quality sequence stops: 1
      Source: IMAGE Consortium, LLNL
      This clone is available royalty-free through LLNL; contact the
      IMAGE Consortium (info@image.llnl.gov) for further information.
      Trace considered overall poor quality
      Insert Length: 1054 Std Error: 0.00
      Seq primer: Promega -21ml3
      High quality sequence stop: 1.
      Location/Qualifiers
        1. .49
          /organism="Homo sapiens"
          /db_xref="GDB:421391"

FEATURES
  source
    Location/Qualifiers
      1. .31
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:2221832"
        /clone_lib="NCI_CGAP_Pan1"
        /tissue_type="adenocarcinoma"
        /lab_host="DH10B"
        /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
        Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
        Average insert size 1.72 kb. Life Technologies catalog #:
        11548-013"
      8 a 7 c 8 g 8 t
      BASE COUNT 8 a 7 c 8 g 8 t
  
```

Qy 1 ccttgaagggaattt 14

Search completed: May 23, 2000, 11:00:17  
Job time: 5899 sec

Search completed: May 23, 2000, 11:00:17  
Job time: 5899 sec

• • • • •

Tue May 23 16:13:53 2000

us-08-945-805-1.lim50.rst

Page 9

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:40:43 ; Search time 68.55 seconds  
(without alignments)  
37.924 Million cell updates/sec

Title: US-08-945-805-1

Perfect score: 20

Sequence: 1 ccttgaaggattccctcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 328916

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14.2	71.0	21	3	US-08-410-654B-34
2	14.2	71.0	21	3	US-08-474-851-34
3	14.2	71.0	21	3	US-08-481-560-34
4	14.2	71.0	26	3	US-08-412-376-31
5	13.2	66.0	27	1	US-08-248-021A-25
6	13.2	66.0	36	3	US-08-412-376-30
7	13	65.0	35	4	US-08-920-165-1
8	12.8	64.0	20	3	US-08-117-952-411
9	12.8	64.0	29	1	US-08-333-565-46
10	12.8	64.0	29	3	US-08-661-479-46
11	12.6	63.0	22	5	US-08-594-452-70
12	12.6	63.0	33	3	US-08-412-376-25
13	12.6	63.0	33	3	US-08-412-376-25
14	12.6	63.0	43	3	US-08-179-557-3
15	12.4	62.0	26	1	US-08-354-101-7
16	12.4	62.0	26	2	US-08-470-834-7
17	12.2	61.0	20	2	US-08-462-305-14
18	12.2	61.0	20	3	US-08-613-417A-14
19	12.2	61.0	20	5	US-08-594-452-14
20	12.2	61.0	20	5	US-08-578-686C-13
21	12.2	61.0	20	5	US-08-281-203-16
22	12.2	61.0	20	5	US-08-594-452-69
23	12.2	61.0	26	4	US-08-859-998-1357
24	12.2	61.0	31	4	US-08-602-359A-11
25	12.2	61.0	39	4	US-08-857-946-58
26	12.2	61.0	39	5	US-08-970-740-58
27	12	60.0	32	5	US-08-705-245-35

Sequence 24, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 12, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 2, Appl  
Sequence 3, Appl  
Sequence 5, Appl  
Patent No. 5194428  
Sequence 11, Appl  
Sequence 37, Appl  
Sequence 8, Appl  
Sequence 20, Appl  
Sequence 20, Appl  
Sequence 2, Appl  
Sequence 426, App  
Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-08-410-654B-34  
; Sequence 34, Application US/08410654B  
; Patent No. 5833976  
; GENERAL INFORMATION:  
; APPLICANT: Rene de Waal Malefyt  
; APPLICANT: Di-Hwei Hsu  
; APPLICANT: Anne O'Garra  
; APPLICANT: Hergen Spits  
; TITLE OF INVENTION: Use of Interleukin-10 to Treat  
; TITLE OF INVENTION: Septic Shock  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schering-Plough Corporation  
; STREET: 2000 Galloping Hill Road  
; CITY: Kenilworth  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07033  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: 7.5.3  
; SOFTWARE: Microsoft Word 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/410,654B  
; FILING DATE: 24-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/229,854  
; FILING DATE: 19-APR-1994  
; APPLICATION NUMBER: US 07/926,853  
; FILING DATE: 06-AUG-1992  
; APPLICATION NUMBER: US 07/742,129  
; FILING DATE: 06-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Foulke, Cynthia L.  
; REGISTRATION NUMBER: 32,364  
; REFERENCE/DOCKET NUMBER: DX0221KQ1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-298-2987  
; TELEFAX: 908-298-5388  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (oligonucleotide)  
US-08-410-654B-34

Query Match 71.0%; Score 14.2; DB 3; Length 21;  
Best Local Similarity 84.2%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctc 19  
||| ||||| |||||  
Db 3 CCTGGAAGGATCTCCCC 21

## RESULT 2

US-08-474-851-34  
; Sequence 34, Application US/08474851  
; Patent No. 5837232  
; GENERAL INFORMATION:  
; APPLICANT: Rene de Waal Malefyt  
; APPLICANT: Di-Hwei Hsu  
; APPLICANT: Anne O'Garra  
; APPLICANT: Hergen Spits  
; TITLE OF INVENTION: Use of An Interleukin-10 Antagonist to Treat  
; TITLE OF INVENTION: A B Cell Mediated Autoimmune Disorder  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schering-plough Corporation  
; STREET: 2000 Galloping Hill Road  
; CITY: Kenilworth  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07033

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: 7.5.3  
SOFTWARE: Microsoft Word 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,851  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/410,654  
FILING DATE: 24-MAR-1995  
APPLICATION NUMBER: US 08/229,854  
FILING DATE: 19-APR-1994  
APPLICATION NUMBER: US 07/926,853  
FILING DATE: 06-AUG-1992  
APPLICATION NUMBER: US 07/742,129  
FILING DATE: 06-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Foulke, Cynthia L.  
REGISTRATION NUMBER: 32,364  
REFERENCE/DOCKET NUMBER: DX0221KQ1GD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-298-5388  
TELEFAX: 908-298-5388  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (oligonucleotide)  
US-08-474-851-34

Query Match 71.0%; Score 14.2; DB 3; Length 21;  
Best Local Similarity 84.2%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctc 19  
||| ||||| |||||  
Db 3 CCTGGAAGGATCTCCCC 21

## RESULT 3

US-08-481-560-34  
; Sequence 34, Application US/08481560  
; Patent No. 5837293  
; GENERAL INFORMATION:  
; APPLICANT: Rene de Waal Malefyt  
; APPLICANT: Di-Hwei Hsu  
; APPLICANT: Anne O'Garra  
; APPLICANT: Hergen Spits  
; TITLE OF INVENTION: Use of Interleukin-10 to Modulate  
; TITLE OF INVENTION: Inflammation or T-Cell Mediated  
; TITLE OF INVENTION: Immune Function  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schering-Plough Corporation  
; STREET: 2000 Galloping Hill Road  
; CITY: Kenilworth  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07033

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: 7.5.3  
SOFTWARE: Microsoft Word 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,560  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/410,654  
FILING DATE: 24-MAR-1995  
APPLICATION NUMBER: US 08/229,854  
FILING DATE: 19-APR-1994  
APPLICATION NUMBER: US 07/926,853  
FILING DATE: 06-AUG-1992  
APPLICATION NUMBER: US 07/742,129  
FILING DATE: 06-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Foulke, Cynthia L.  
REGISTRATION NUMBER: 32,364  
REFERENCE/DOCKET NUMBER: DX0221KQ1GC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-298-2987  
TELEFAX: 908-298-5388  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (oligonucleotide)  
US-08-481-560-34

Query Match 71.0%; Score 14.2; DB 3; Length 21;  
Best Local Similarity 84.2%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctc 19  
||| ||||| |||||  
Db 3 CCTGGAAGGATCTCCCC 21

## RESULT 4

US-08-412-376-31  
; Sequence 31, Application US/08412376  
; Patent No. 5849900  
; GENERAL INFORMATION:  
; APPLICANT: Moelling, Karin  
; TITLE OF INVENTION: Inhibition Of Viruses By Antisense  
; TITLE OF INVENTION: Oligomers Capable Of Binding To Polypurine-Rich Tract Of S1

;  
; TITLE OF INVENTION: Stranded RNA Or RNA-DNA Hybrids  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz  
; ADDRESSEE: Mackiewicz & No. 5849900ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/412,376  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/954,184  
; FILING DATE: 29-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Doreen Yatko Trujillo  
; REGISTRATION NUMBER: 35,719  
; REFERENCE/DOCKET NUMBER: APOL-0021  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; ANTI-SENSE: Yes  
; US-08-412-376-31

Query Match 71.0%; Score 14.2; DB 3; Length 36;  
Best Local Similarity 84.2%; Pred. No. 28;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cttgaaggattccctcc 20  
|| ||||| ||||| |  
Db 8 CTGGAAGGTTTCCCTC 26

RESULT 5  
US-08-248-021A-25/c  
; Sequence 25, Application US/08248021A  
; Patent No. 5648240  
; GENERAL INFORMATION:  
; APPLICANT: Hook, Magnus  
; APPLICANT: Jonsson, Klas  
; APPLICANT: Patti, Joseph M.  
; APPLICANT: Gurusiddappa, Sivashankarappa  
; TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States of America  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/248,021A

;  
; FILING DATE: 24-MAY-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: TARK:155  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-248-021A-25

Query Match 66.0%; Score 13.2; DB 1; Length 27;  
Best Local Similarity 83.3%; Pred. No. 91;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cttgaaggattccctc 19  
|| ||||| ||||| |  
Db 26 CTTGAAGTGATTGCCG 9

RESULT 6  
US-08-412-376-30  
; Sequence 30, Application US/08412376  
; Patent No. 5849900  
; GENERAL INFORMATION:  
; APPLICANT: Moelling, Karin  
; TITLE OF INVENTION: Inhibition Of Viruses By Antisense  
; TITLE OF INVENTION: Oligomers Capable Of Binding To Polypurine-Rich Tract Of Si  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz  
; ADDRESSEE: Mackiewicz & No. 5849900ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/412,376  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/954,184  
; FILING DATE: 29-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Doreen Yatko Trujillo  
; REGISTRATION NUMBER: 35,719  
; REFERENCE/DOCKET NUMBER: APOL-0021  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; ANTI-SENSE: Yes  
; US-08-412-376-30

Query Match 66.0%; Score 13.2; DB 3; Length 36;  
Best Local Similarity 83.3%; Pred. No. 97;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ttgaaggatttcctcc 20  
Db 9 TGAAGGTTTCCCTC 26

RESULT 7  
US-08-920-165-1  
; Sequence 1, Application US/08920165  
; Patent No. 5972608  
; GENERAL INFORMATION:  
; APPLICANT: Peterson, Craig L.  
; APPLICANT: Logie, Colin  
; TITLE OF INVENTION: ASSAYS AND REAGENTS FOR CHROMATIN  
; TITLE OF INVENTION: REMODELING ENZYMES AND THEIR MODULATORS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/920,165  
FILING DATE: 27-AUG-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07917/057001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Other  
US-08-920-163-1

Query Match 65.0%; Score 13; DB 4; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cttgaaggattt 14  
Db 20 CTTGAAGGATTT 32

RESULT 8  
US-08-117-952-411  
; Sequence 411, Application US/08117952  
; Patent No. 5851760  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; APPLICANT: Smith, Michael W.  
; TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE  
; TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES  
; NUMBER OF SEQUENCES: 797

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90071

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/117,952  
FILING DATE: 07-SEP-1993  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/078,471  
FILING DATE: 15-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9423  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 411:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Oligonucleotide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-117-952-411

Query Match 64.0%; Score 12.8; DB 3; Length 20;  
Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tgaaggatttcctc 19  
Db 2 TGAGGGCATTTCCCTC 17

RESULT 9  
US-08-333-565-46  
; Sequence 46, Application US/08333565  
; Patent No. 5622852  
; GENERAL INFORMATION:  
; APPLICANT: KORSMEYER, Stanley J.  
; TITLE OF INVENTION: Bcl-x/Bcl-2 ASSOCIATED CELL DEATH  
; TITLE OF INVENTION: REGULATOR  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourile and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/333,565  
FILING DATE: 31-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 15726A-000700  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (oligo)  
US-08-333-565-46

Query Match 64.0%; Score 12.8; DB 1; Length 29;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccttgaaggatttcc 16  
||||| ||||| |||||  
Db 12 CCTTCAGGACTTCC 27

RESULT 10  
US-08-661-479-46  
Sequence 46, Application US/08661479  
Patent No. 5834209  
GENERAL INFORMATION:  
APPLICANT: KORSMEYER, Stanley J.  
TITLE OF INVENTION: Bcl-X/Bcl-2 ASSOCIATED CELL DEATH  
TITLE OF INVENTION: REGULATOR  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/661.479  
FILING DATE: 11-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/333,565  
FILING DATE: 31-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 15726A-000700  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (oligo)  
US-08-661-479-46

Query Match 64.0%; Score 12.8; DB 3; Length 29;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccttgaaggatttcc 16  
||||| ||||| |||||  
Db 12 CCTTCAGGACTTCC 27

RESULT 11  
US-08-594-452-70/c  
Sequence 70, Application US/08594452  
Patent No. 6013639  
GENERAL INFORMATION:  
APPLICANT: PEYMAN, Anushirwan  
APPLICANT: UHLMANN, Eugen  
TITLE OF INVENTION: G CAP-STABILIZED OLIGONUCLEOTIDES  
NUMBER OF SEQUENCES: 105  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/594,452  
FILING DATE: 31-JAN-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 195 02 912.7  
FILING DATE: 31-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: SANDERCOCK, Colin G.  
REGISTRATION NUMBER: 31,298  
REFERENCE/DOCKET NUMBER: 18748/264/HOCE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-594-452-70

Query Match 63.0%; Score 12.6; DB 5; Length 22;  
Best Local Similarity 78.9%; Pred. No. 1.8e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccttgaaggatttccctc 19  
||||| ||||| ||||| |  
Db 19 CCTTGCCGGGCTTCCCCC 1

RESULT 12  
US-08-412-376-25  
Sequence 25, Application US/08412376  
Patent No. 5849900  
GENERAL INFORMATION:  
APPLICANT: Moelling, Karin  
TITLE OF INVENTION: Inhibition Of Viruses By Antisense  
TITLE OF INVENTION: Oligomers Capable Of Binding To Polypurine-Rich Tract Of Si  
TITLE OF INVENTION: Stranded RNA Or RNA-DNA Hybrids  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz  
ADDRESSEE: Mackiewicz & No. 5849900ris

STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
FILING DATE: Herewith  
APPLICATION NUMBER: US/08/412,376  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/954,184  
FILING DATE: 29-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Doreen Yatko Trujillo  
REGISTRATION NUMBER: 35,719  
REFERENCE/DOCKET NUMBER: APOL-0021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
ANTI-SENSE: Yes  
US-08-412-376-25

Query Match 63.0%; Score 12.6; DB 3; Length 33;  
Best Local Similarity 78.9%; Pred. No. 2e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cttgaaggatttcctcc 20  
||| ||||| ||||| |  
Db 6 CTGGGAGGGTTTCCCTTC 24

RESULT 13  
US-08-412-376-25/c  
Sequence 25, Application US/08412376  
Patent No. 5849900  
GENERAL INFORMATION:  
APPLICANT: Moelling, Karin  
TITLE OF INVENTION: Inhibition Of Viruses By Antisense  
TITLE OF INVENTION: Oligomers Capable Of Binding To Polypurine-Rich Tract Of Singl  
TITLE OF INVENTION: Stranded RNA Or RNA-DNA Hybrids  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz  
ADDRESSEE: Mackiewicz & No. 5849900ris  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/412,376  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/954,184  
FILING DATE: 29-SEP-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Doreen Yatko Trujillo  
REGISTRATION NUMBER: 35,719  
REFERENCE/DOCKET NUMBER: APOL-0021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
ANTI-SENSE: Yes  
US-08-412-376-25

Query Match 63.0%; Score 12.6; DB 3; Length 33;  
Best Local Similarity 78.9%; Pred. No. 2e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cttgaaggatttcctcc 20  
||| ||||| ||||| |  
Db 27 CTGGAAGGGAACCTCC 9

RESULT 14  
US-08-179-557-3  
Sequence 3, Application US/08179557  
Patent No. 5837509  
GENERAL INFORMATION:  
APPLICANT: ISRAELSEN, Hans  
APPLICANT: BECH HANSEN, Egon  
APPLICANT: MADSEN, Soeren Michael  
APPLICANT: JOHANSEN, Eric  
APPLICANT: NILSSON, Dan  
APPLICANT: VRANG, Astrid  
TITLE OF INVENTION: Recombinant Lactic Acid Bacterium  
TITLE OF INVENTION: Containing an Inserted Promoter and Method of Constructing  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/179,557  
FILING DATE: 07-JAN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 1579/92  
FILING DATE: 30-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 0988/93  
FILING DATE: 01-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK94/00004  
FILING DATE: 03-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,681  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30307/140/PLVI  
TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202 672 5300  
; TELEFAX: 202 672 5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 43 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-179-557-3

Query Match 63.0%; Score 12.6; DB 3; Length 43;  
Best Local Similarity 78.9%; Pred. No. 2.1e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccttgaagggtttccctc 19  
| |||| |||| ||||  
Db 25 CTTTGAAGGATATTCCTC 43

RESULT 15  
US-08-354-101-7  
; Sequence 7, Application US/08354101  
; Patent No. 5665591  
; GENERAL INFORMATION:  
; APPLICANT: Sonenshein, Gail E.  
; APPLICANT: Lawrence, Roger  
; APPLICANT: Bellas, Robert E.  
; TITLE OF INVENTION: Regulation of Smooth Muscle Cell  
; TITLE OF INVENTION: Proliferation  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Baker & Botts, L.L.P.  
; STREET: The Warner; 1299 Pennsylvania Ave., N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004-2400  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/354,101  
; FILING DATE: 06-DEC-1994  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Remenick, James  
; REGISTRATION NUMBER: 36,902  
; REFERENCE/DOCKET NUMBER: 16865-0134  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-639-7700  
; TELEFAX: 202-639-7890  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-354-101-7

Query Match 62.0%; Score 12.4; DB 1; Length 26;  
Best Local Similarity 92.9%; Pred. No. 2.4e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 tgaagggtttccc 17  
| ||||| |||||

Db 8 TGAAGGGTTTTCCC 21

Search completed: May 23, 2000, 11:19:44  
Job time: 5941 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 10:36:29 ; Search time 236.64 Seconds  
(without alignments)  
-82.217 Million cell updates/sec

US-08-945-805-2

Perfect score: 20  
Sequence: 1 ttgcgcgtacctgacttagcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 285916

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_bal.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15.8	79.0	39	5	AR009260	Sequence 9
2	13.6	68.0	22	5	A51393	Sequence 9
3	12.6	63.0	24	5	AR066517	Sequence 9
4	12.6	63.0	33	5	AR009600	Sequence 9
5	12.6	63.0	33	5	I44060	Sequence 9
6	12.6	63.0	39	5	A36980	Sequence 31
c 7	12.6	63.0	39	5	A36987	Sequence 31
8	12.6	63.0	39	5	A39188	Sequence 36
c 9	12.6	63.0	39	5	A39195	Sequence 43
10	12.6	63.0	39	5	A39247	Sequence 31
c 11	12.6	63.0	39	5	A39254	Sequence 38
c 12	12.4	62.0	32	5	AR053018	Sequence 38
13	12	60.0	30	5	AR028275	Sequence 38
c 14	12	60.0	36	5	A92661	Sequence 9
c 15	12	60.0	37	5	A07461	Nucleotide
c 16	12	60.0	37	5	A07462	Nucleotide
c 17	12	60.0	37	5	A07463	Nucleotide
c 18	12	60.0	50	40	AF057506	Homo sapi
c 19	11.8	59.0	20	5	AR055401	Sequence
20	11.8	59.0	30	5	AR031195	Sequence
21	11.8	59.0	30	5	I17196	Sequence 4
22	11.8	59.0	30	5	I75967	Sequence 4
c 23	11.6	58.0	21	5	AR066799	Sequence 30
c 24	11.6	58.0	39	5	A84043	Sequence 30
c 25	11.6	58.0	39	5	A84044	Sequence 31
c 26	11.6	58.0	43	5	AR014640	Sequence 31
27	11.6	58.0	43	5	AR014641	Sequence 31
c 28	11.4	57.0	19	5	A59325	Sequence 12
c 29	11.4	57.0	20	5	A30034	Oligonucleo
c 30	11.4	57.0	22	5	A62812	Sequence 53
c 31	11.2	56.0	19	5	AR050710	Sequence
32	11.2	56.0	21	5	A51144	Sequence 13
33	11.2	56.0	21	5	A76969	Sequence 13
34	11.2	56.0	22	13	DOG00203A	Canis fami
35	11.2	56.0	23	5	A92601	Sequence 7
c 36	11.2	56.0	40	5	E07396	PCR primer
37	11.2	56.0	40	5	E07397	PCR primer
c 38	11	55.0	23	5	AR001099	Sequence
c 39	11	55.0	23	5	AR027465	Sequence
c 40	11	55.0	23	5	AR058274	Sequence
c 41	11	55.0	23	5	I18531	Sequence 21
c 42	11	55.0	23	5	I66631	Sequence 21
c 43	11	55.0	23	5	I88806	Sequence 21
c 44	11	55.0	24	5	I64405	Sequence 21
c 45	11	55.0	33	5	AR033935	Sequence

ALIGNMENTS

```
RESULT 1
AR009260 LOCUS AR009260 39 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 23 from patent US 5756291.
ACCESSION AR009260
VERSION AR009260.1 GI:3968065
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 39)
AUTHORS Griffin,L., Albrecht,G., Latham,J., Leung,L., Vermaas,E. and Toole,J.J.
TITLE Aptamers specific for biomolecules and methods of making
JOURNAL Patent: US 5756291-A 23 26-MAY-1998;
FEATURES
source Location/Qualifiers
1..39
/organism="unknown"
BASE COUNT 11 a 11 c 7 g 9 t 1 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 5; Length 39;
Best Local Similarity 85.0%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20
||||| ||||| |||
Db 14 TTGCCGNACCTGAATTCGCC 33

RESULT 2
AR009600 LOCUS AR009600 22 bp DNA PAT 10-MAR-1997
DEFINITION Sequence 9 from Patent WO9617071.
ACCESSION AR009600
VERSION AR009600.1 GI:2304212
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 22)
AUTHORS Cohen-Haguenauer,O.
TITLE ENCAPSIDATION CELL LINES FOR THE TRANSCOMPLEMENTATION OF DEFECTIVE RETROVIRAL VECTORS
JOURNAL Patent: WO 9617071-A 9 06-JUN-1996;
COMMENT COHEN HAGUENAUER ODILE (FR)
Other publication AU 4306996 960619
Other publication FR 2727429 960531.
FEATURES
source Location/Qualifiers
1..22
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 6 a 6 c 5 g 5 t
ORIGIN

Query Match 68.0%; Score 13.6; DB 5; Length 22;
Best Local Similarity 80.0%; Pred. No. 7.1e+03;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20
||||| || ||||| ||||
Db 3 TTGCCGACTTGACTAAGCC 22

RESULT 3
AR066517 LOCUS AR066517 24 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5851521.
ACCESSION AR066517
VERSION AR066517.1 GI:5997739

KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 33)
AUTHORS Baszczyński,C., Barbour,E., Miki,B. and Hattori,J.
TITLE Microspore-specific regulatory element
JOURNAL Patent: US 5633438-A 4 27-MAY-1997;
FEATURES
source Location/Qualifiers
1..33
/organism="unknown"
```

---

```
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Branellec,D., Walsh,K., Isner,J.M. and Benefle,P.
TITLE Viral vectors and their use for treating hyperproliferative disorders, in particular restenosis
JOURNAL Patent: US 5851521-A 2 22-DEC-1998;
FEATURES
source Location/Qualifiers
1..24
/organism="unknown"
BASE COUNT 4 a 8 c 5 g 7 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 24;
Best Local Similarity 78.9%; Pred. No. 2.4e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttagcc 20
||||| ||||| |||
Db 6 TGACGTGCTGACTATGCC 24

RESULT 4
AR009600 LOCUS AR009600 33 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 4 from patent US 5756324.
ACCESSION AR009600
VERSION AR009600.1 GI:3968405
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 33)
AUTHORS Baszczyński,C., Barbour,E., Miki,B. and Hattori,J.
TITLE Use of a microspore-specific regulatory element for production of virus and insect-resistant plants
JOURNAL Patent: US 5756324-A 4 26-MAY-1998;
FEATURES
source Location/Qualifiers
1..33
/organism="unknown"
BASE COUNT 5 a 14 c 7 g 7 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 33;
Best Local Similarity 78.9%; Pred. No. 2.4e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttagcc 20
||||| ||||| |||||
Db 12 TGCCTTACCGCTCTCAGCC 30

RESULT 5
AR009600 LOCUS AR009600 33 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 4 from patent US 5633438.
ACCESSION AR009600
VERSION AR009600.1 GI:2469158
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 33)
AUTHORS Baszczyński,C., Barbour,E., Miki,B. and Hattori,J.
TITLE Microspore-specific regulatory element
JOURNAL Patent: US 5633438-A 4 27-MAY-1997;
FEATURES
source Location/Qualifiers
1..33
/organism="unknown"
```

	Other publication AU 4700793 940303
	Other publication ZA 9305530 950130
	Other publication JP 8502084T 960305.
FEATURES	Location/Qualifiers
source	1..39
/organism=	"unidentified"
/db_xref=	"taxon:32644"
BASE COUNT	13 a 11 c 9 g 6 t
ORIGIN	
Query Match	63.0%; Score 12.6; DB 5; Length 39;
Best Local Similarity	78.9%; Pred. No. 2.4e+04;
Matches	15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 ttgcctacgtgacttagc 19 
Db	26 TTGCTGCACCTCCTACTGGC 8
RESULT	8
A39188	
LOCUS	A39188 39 bp DNA PAT 05-MAR-1997
DEFINITION	Sequence 36 from Patent WO9414963.
ACCESSION	A39188
VERSION	A39188.1 GI:2295557
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 39)
AUTHORS	Egmond,M.R., Van,D.H., Musters,W., Peters,H., Verrips,C.T. and De,V.J.
TITLE	MODIFIED CUTINASES, DNA, VECTOR AND HOST
JOURNAL	Patent: WO 9414963-A 36 07-JUL-1994;
COMMENT	UNILEVER PLC (GB) Other publication SK 79595 951108 Other publication PL 309388 951002 Other publication CA 2150837 940707 Other publication AU 569994 940719 Other publication CN 1090328 940803 Other publication CZ 9501578 951213 Other publication ZA 9309415 950615. Location/Qualifiers
FEATURES	Location/Qualifiers
source	1..39
/organism=	"unidentified"
/db_xref=	"taxon:32644"
BASE COUNT	6 a 12 c 9 g 12 t
ORIGIN	
Query Match	63.0%; Score 12.6; DB 5; Length 39;
Best Local Similarity	78.9%; Pred. No. 2.4e+04;
Matches	15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 ttgcctacgtgacttagc 19 
Db	5 TTGCTGCACCTCCTACTGGC 23
RESULT	9
A39195/c	
LOCUS	A39195 39 bp DNA PAT 05-MAR-1997
DEFINITION	Sequence 43 from Patent WO9414963.
ACCESSION	A39195
VERSION	A39195.1 GI:2295564
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 39)
AUTHORS	Egmond,M.R., Van,D.H., Musters,W., Peters,H., Verrips,C.T. and De,V.J.
TITLE	ENZYMATIC DETERGENT COMPOSITIONS
JOURNAL	Patent: WO 9403578-A 38 17-FEB-1994;
COMMENT	UNILEVER PLC (GB) Other publication CZ 9500242 951018 Other publication CN 1088256 940622 Other publication PL 307269 950515 Other publication AU 4700793 940303 Other publication ZA 9305530 950130 Other publication JP 8502084T 960305. Location/Qualifiers
FEATURES	Location/Qualifiers
source	1..39
/organism=	"unidentified"
/db_xref=	"taxon:32644"
BASE COUNT	6 a 12 c 9 g 12 t
ORIGIN	
Query Match	63.0%; Score 12.6; DB 5; Length 39;
Best Local Similarity	78.9%; Pred. No. 2.4e+04;
Matches	15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 ttgcctacgtgacttagc 19 
Db	5 TTGCTGCACCTCCTACTGGC 23
RESULT	7
A36987/c	
LOCUS	A36987 39 bp DNA PAT 05-MAR-1997
DEFINITION	Sequence 38 from Patent WO9403578.
ACCESSION	A36987
VERSION	A36987.1 GI:2294181
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 39)
AUTHORS	Hijden,H.T., Marugg,J.D., Warr,J.F., Klugkist,J., Musters,W. and Hondmann,D.H.
TITLE	ENZYMATIC DETERGENT COMPOSITIONS
JOURNAL	Patent: WO 9403578-A 38 17-FEB-1994;
COMMENT	UNILEVER PLC (GB) Other publication CZ 9500242 951018 Other publication CN 1088256 940622 Other publication PL 307269 950515 Other publication AU 4700793 940303 Other publication ZA 9305530 950130 Other publication JP 8502084T 960305. Location/Qualifiers
FEATURES	Location/Qualifiers
source	1..39
/organism=	"unidentified"
/db_xref=	"taxon:32644"
BASE COUNT	6 a 12 c 9 g 12 t
ORIGIN	
Query Match	63.0%; Score 12.6; DB 5; Length 33;
Best Local Similarity	78.9%; Pred. No. 2.4e+04;
Matches	15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	2 tgccgtacctgacttagc 20 
Db	12 TGCCTTACCGTCCTCAGCC 30
RESULT	6
A36980	
LOCUS	A36980 39 bp DNA PAT 05-MAR-1997
DEFINITION	Sequence 31 from Patent WO9403578.
ACCESSION	A36980
VERSION	A36980.1 GI:2294174
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 39)
AUTHORS	Hijden,H.T., Marugg,J.D., Warr,J.F., Klugkist,J., Musters,W. and Hondmann,D.H.
TITLE	ENZYMATIC DETERGENT COMPOSITIONS
JOURNAL	Patent: WO 9403578-A 31 17-FEB-1994;
COMMENT	UNILEVER PLC (GB) Other publication CZ 9500242 951018 Other publication CN 1088256 940622 Other publication PL 307269 950515 Other publication AU 4700793 940217 Other publication AU 4700793 940303 Other publication ZA 9305530 950130 Other publication JP 8502084T 960305. Location/Qualifiers
FEATURES	Location/Qualifiers
source	1..39
/organism=	"unidentified"
/db_xref=	"taxon:32644"
BASE COUNT	6 a 12 c 9 g 12 t
ORIGIN	
Query Match	63.0%; Score 12.6; DB 5; Length 33;
Best Local Similarity	78.9%; Pred. No. 2.4e+04;
Matches	15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 ttgcctacgtgacttagc 19 
Db	5 TTGCTGCACCTCCTACTGGC 23
RESULT	7
A36987/c	
LOCUS	A36987 39 bp DNA PAT 05-MAR-1997
DEFINITION	Sequence 38 from Patent WO9403578.
ACCESSION	A36987
VERSION	A36987.1 GI:2294181
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 39)
AUTHORS	Hijden,H.T., Marugg,J.D., Warr,J.F., Klugkist,J., Musters,W. and Hondmann,D.H.
TITLE	ENZYMATIC DETERGENT COMPOSITIONS
JOURNAL	Patent: WO 9403578-A 38 17-FEB-1994;
COMMENT	UNILEVER PLC (GB) Other publication CZ 9500242 951018 Other publication CN 1088256 940622 Other publication PL 307269 950515 Other publication AU 4700793 940217 Other publication ZA 9305530 950130 Other publication JP 8502084T

TITLE  
JOURNAL  
COMMENT  
MODIFIED CUTINASES, DNA, VECTOR AND HOST  
Patent: WO 9414963-A 43 07-JUL-1994;  
UNILEVER PLC (GB)  
Other publication SK 79595 951108  
Other publication PL 309388 951002  
Other publication CA 2150837 940707  
Other publication AU 5699994 940719  
Other publication CN 1090328 940803  
Other publication CZ 9501578 951213  
Other publication ZA 9309415 950615.  
Location/Qualifiers  
1. .39  
/organism="unidentified"  
/db\_xref="taxon:32644" 6 t  
13 a 11 c 9 g 6 t  
BASE COUNT  
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 39;  
Best Local Similarity 78.9%; Pred. No. 2.4e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgcgtacctgacttagc 19  
||||| ||||| ||||| ||  
Db 26 TTGCTGCACCTCACTTGGC 8

RESULT 10  
A39247  
LOCUS A39247 39 bp DNA PAT 05-MAR-1997  
DEFINITION Sequence 31 from Patent WO9414964.  
ACCESSION A39247  
VERSION A39247.1 GI:2295612  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS Egmond,M.R., Van,D.H., Musters,W., Peters,H., Verrrips,C.T. and De,V.J.  
TITLE MODIFIED CUTINASES, DNA, VECTOR AND HOST  
JOURNAL Patent: WO 9414964-A 31 07-JUL-1994;  
UNILEVER NV (NL)  
Other publication CA 2151038 940707  
Other publication HU 71315 951128  
Other publication PL 309403 951002  
Other publication AU 5700094 940719  
Other publication SK 80295 951011  
Other publication CN 1090329 940803  
Other publication CZ 9501639 951115  
Other publication ZA 9309416 950615  
Other publication JP 8504589T 960521.  
Location/Qualifiers  
1. .39  
/organism="unidentified"  
/db\_xref="taxon:32644" 9 g 6 t  
6 a 12 c 9 g 12 t  
BASE COUNT  
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 39;  
Best Local Similarity 78.9%; Pred. No. 2.4e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgcgtacctgacttagc 19  
||||| ||||| ||||| ||  
Db 5 TTGCTGCACCTCACTTGGC 23

RESULT 11  
A39254/c  
LOCUS A39254 39 bp DNA PAT 05-MAR-1997  
DEFINITION Sequence 38 from Patent WO9414964.

ACCESSION A39254  
VERSION A39254.1 GI:2295619  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS Egmond,M.R., Van,D.H., Musters,W., Peters,H., Verrrips,C.T. and De,V.J.  
TITLE MODIFIED CUTINASES, DNA, VECTOR AND HOST  
JOURNAL Patent: WO 9414964-A 38 07-JUL-1994;  
UNILEVER NV (NL)  
Other publication CA 2151038 940707  
Other publication HU 71315 951128  
Other publication PL 309403 951002  
Other publication AU 5700094 940719  
Other publication SK 80295 951011  
Other publication CN 1090329 940803  
Other publication CZ 9501639 951115  
Other publication ZA 9309416 950615  
Other publication JP 8504589T 960521.  
Location/Qualifiers  
1. .39  
/organism="unidentified"  
/db\_xref="taxon:32644" 11 c 9 g 6 t  
13 a 11 c 9 g 6 t  
BASE COUNT  
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 39;  
Best Local Similarity 78.9%; Pred. No. 2.4e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgcgtacctgacttagc 19  
||||| ||||| ||||| ||  
Db 26 TTGCTGCACCTCACTTGGC 8

RESULT 12  
AR053018/c  
LOCUS AR053018 32 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 62 from patent US 5833991.  
ACCESSION AR053018  
VERSION AR053018.1 GI:5977880  
KEYWORDS  
SOURCE  
ORGANISM  
Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Masucci,M.G.  
TITLE Glycine-containing sequences conferring invisibility to the immune system  
JOURNAL Patent: US 5833991-A 62 10-NOV-1998;  
DEFINITION Location/Qualifiers  
1. .32  
/organism="unknown"  
9 a 7 c 5 g 11 t  
BASE COUNT  
ORIGIN

Query Match 62.0%; Score 12.4; DB 5; Length 32;  
Best Local Similarity 92.9%; Pred. No. 3.1e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 cgtacctgacttag 18  
||||| ||||| ||||| ||  
Db 28 CGTAACTGACTTAG 15

RESULT 13  
AR028275  
LOCUS AR028275 30 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 124 from patent US 5858661.

ACCESSION AR028275  
VERSION AR028275.1 GI:5940248  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Shiloh,Y.  
TITLE Ataxia-telangiectasia gene and its genomic organization  
JOURNAL Patent: US 5858661-A 124 12-JAN-1999;  
FEATURES Location/Qualifiers  
source 1..30  
/organism="unknown"  
BASE COUNT 2 a 11 c 6 g 11 t  
ORIGIN

Query Match 60.0%; Score 12; DB 5; Length 30;  
Best Local Similarity 75.0%; Pred. No. 5.le+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20  
||||| ||||| |||||  
Db 8 TTCCCTTCTCTGGCTTAGCC 27

RESULT 14  
A92661/c  
LOCUS A92661 36 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 9 from Patent WO9806831.  
ACCESSION A92661  
VERSION A92661.1 GI:6741301  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 36)  
AUTHORS La,C.U. and Willmitzer,L.  
TITLE TRANSGENIC PLANT CELLS AND PLANTS WITH MODIFIED ACETYL-COA  
FORMATION  
JOURNAL Patent: WO 9806831-A 19-FEB-1998;  
MAX PLANCK GESELLSCHAFT (DE); WILLMITZER LOTHAR (DE)

FEATURES Location/Qualifiers  
source 1..36  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 12 a 9 c 8 g 7 t  
ORIGIN

Query Match 60.0%; Score 12; DB 5; Length 36;  
Best Local Similarity 75.0%; Pred. No. 5.e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20  
||||| ||||| |||||  
Db 26 TTGCCTTGGCTGACATAGCC 7

RESULT 15  
A07461/c  
LOCUS A07461 37 bp DNA PAT 21-JUL-1993  
DEFINITION Nucleotide sequence 5 from patent number EP0389347.  
ACCESSION A07461  
VERSION A07461.1 GI:411445  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 37)  
AUTHORS Audonnet,J.C. and Bruneau,P.  
TITLE Vaccines against septicemic bacteria  
JOURNAL Patent: EP 0389347-A 5 26-SEP-1990;

FEATURES RHONE MERIEUX  
Location/Qualifiers  
source 1..37  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 7 a 12 c 8 g 10 t  
ORIGIN

Query Match 60.0%; Score 12; DB 5; Length 37;  
Best Local Similarity 75.0%; Pred. No. 5.e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20  
||||| ||||| |||||  
Db 30 TTGTCGTAATCGACTGAGCC 11

Search completed: May 23, 2000, 10:36:32  
Job time: 5933 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 11:21:29 ; Search time 96.59 seconds  
(without alignments)  
51.805 Million cell updates/sec

Title: US-08-945-805-2

Perfect score: 20

Sequence: 1 ttgccgtacctgacttagcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 391736

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	1	V59191
2	13.6	68.0	22	1	T31191
3	12.4	62.0	32	1	V02751
4	12	60.0	20	1	T96018
5	12	60.0	36	1	V11387
6	11.8	59.0	17	1	V93645
7	11.8	59.0	20	1	T95409
8	11.8	59.0	21	1	Q37759
9	11.8	59.0	27	1	V70405
10	11.8	59.0	30	1	Q78691
11	11.8	59.0	30	1	T10549
12	11.8	59.0	30	1	T86542
13	11.8	59.0	31	1	V21467
14	11.6	58.0	14	1	N82409
15	11.6	58.0	21	1	Q82147
16	11.6	58.0	23	1	Q25902
17	11.6	58.0	23	1	T59620
18	11.6	58.0	39	1	Q67030
19	11.6	58.0	41	1	V50845
20	11.6	58.0	43	1	Q49575
21	11.6	58.0	43	1	Q49576
22	11.6	58.0	43	1	Q95015
23	11.6	58.0	43	1	Q95016
24	11.6	58.0	43	1	T11018
25	11.4	57.0	18	1	V61420
26	11.4	57.0	19	1	T87812
27	11.4	57.0	22	1	T92484
28	11.4	57.0	33	1	V15176
29	11.4	57.0	50	1	T48645
30	11.2	56.0	14	1	R50068
31	11.2	56.0	16	1	T69162
32	11.2	56.0	20	1	Q41802
33	11.2	56.0	20	1	Q53938
34	11.2	56.0	21	1	T32459

35 11.2 56.0 23 1 V11650  
36 11.2 56.0 32 1 T93738  
37 11.2 56.0 32 1 T93736  
38 11.2 56.0 33 1 V52102  
39 11.2 56.0 40 1 Q65563  
40 11.2 56.0 40 1 Q65564  
41 11.2 56.0 41 1 V47703  
42 11.2 56.0 45 1 T09248  
43 11 55.0 17 1 T14820  
44 11 55.0 22 1 T35866  
45 11 55.0 23 1 Q72926

E. coli ADSS gene  
PCR primer 2 for a  
PCR primer 2 for a  
Helicobacter poly  
GUS N-terminal/PR1  
Pria signal peptid  
Maize polymorphic  
nACHR alpha4 subun  
Histocyte-secrete  
Glyceraldehyde-3-p  
Forward primer for

#### ALIGNMENTS

RESULT 1

V59191  
ID V59191 standard; DNA; 20 BP.  
AC V59191;  
DT 21-DEC-1998 (first entry)  
DE Random sequence oligonucleotide.  
KW Bak gene; promoter; human; apoptosis; ischaemia; tumour; cancer;  
KW infection; HIV; neurodegenerative disorder; hair loss;  
KW inflammatory bowel disease; transplant rejection; therapy; Bcl-2;  
KW ISRE; interferon-gamma response element; GAS;  
KW gamma interferon-activated sequence; ds.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9835659-A1.  
PD 20-AUG-1998.  
PF 17-FEB-1998; U02845.  
PR 18-FEB-1997; US-038412.  
PA (LXRB-) LAR BIOTECHNOLOGY INC.  
PI Kiefer MC, Ossina N;  
DR WPI; 98-506279/43.  
PT New isolated bak promoter - used for identifying agents that regulate the expression of a gene involved in apoptosis in cells for use in therapy.  
PS Example 11; Page 36; 69pp; English.  
CC This double-stranded oligonucleotide with random sequence was used as a control in experiments to examine regulation of human Bak expression by an ISRE 'decoy' cis-element. Oligonucleotides corresponding to the bak promoter (see V59181) interferon-gamma response element (ISRE, see V59182), gamma-interferon activated sequence (GAS, see V59187) and NF-kappa 1 sequence (see V59189) were used as decoy elements to regulate Bak expression. GAS and NF-kappa B decoy DNAs failed to inhibit Bak expression. The ISRE decoy, but not the random-sequence decoy, markedly inhibited Bak expression in response to interferon-gamma. The ISRE cis-element is critical for Bak gene activation. ISRE decoy DNA can be used as a tool for inhibition of Bak transcription.  
SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T;

Query Match 100.0%; Score 20; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.046;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgccgtacctgacttagcc 20  
|||||  
Db 1 TTGCGGTACTGACTTAGCC 20

RESULT 2

T31191  
ID T31191 standard; DNA; 22 BP.  
AC T31191;  
DT 18-FEB-1997 (first entry)  
DE Friend virus strain FB29 POL sequence PCR primer POL-F.  
KW Retroviral vector; pFOCH29; LTR; long terminal repeat; gag; pol;  
KW encapsidation; transcomplementation; polymerase chain reaction; PCR;  
KW gene therapy; transfer vector; canine cell line; DOGP29; ss.



PT higher fat content in seeds  
 PS Example 8; Page 39; 65pp; German.  
 CC V11379-V11390 are primers used in a method to construct a transgenic  
 CC plant cell with altered acetyl-coenzyme A (AcCoA) hydrolase metabolism  
 CC which has higher activity than wild-type cells. Controlling AcCoA  
 CC hydrolase activity is used to regulate intracellular distribution of  
 CC metabolites, specifically it increases concentration of acetate, leading  
 CC to altered production of isoprenoids, steroids, flavonoids, hormones,  
 CC fats, oils, rubber, alkaloids, sugars and starch and anti-pathogenic  
 CC agents. A specific application is increasing the fat content (by at least  
 CC 7%) of oilseed crops, e.g. (especially) rape, soya, sunflower and oil  
 CC palms, but the method can also be applied to cereals, vegetables,  
 CC sugar beet, tobacco, cotton or ornamental plants.  
 SQ Sequence 36 BP; 12 A; 9 C; 8 G; 7 T;

Query Match 60.0%; Score 12; DB 1; Length 36;  
 Best Local Similarity 75.0%; Pred. No. 8.9e+02; Mismatches 5; Indels 0; Gaps 0;  
 Matches 15; Conservative 0;

Qy 1 ttgcgtacctgacttagcc 20  
 ||||| ||||| |||||  
 Db 26 TTGCTTGGCTGACATAGGC 7

## RESULT 6

V93645/c  
 ID V93645 standard; RNA; 17 BP.

AC V93645;  
 DT 18-FEB-1999 (first entry)  
 DE Human B-raf substrate nucleotide position 2096.  
 KW Human; c-raf; A-raf; B-raf; hammerhead ribozyme; hairpin ribozyme;  
 KW target; substrate; catalyst; modulation; expression; Raf gene;  
 KW delivery; screening; identification; synthesis; deprotection;  
 KW purification; cancer; inflammation; psoriasis; non-hepatic ascites;  
 KW infection; genetic drift; restenosis; rheumatoid arthritis; ss.  
 OS Homo sapiens.  
 PN WO9850530-A2.  
 PD 12-NOV-1998.

PF 05-MAY-1998; U09249.

PR 19-DEC-1997; US-068212.

PR 09-MAY-1997; US-046059.

PR 09-JUN-1997; US-049002.

PR 03-JUL-1997; US-051718.

PR 22-AUG-1997; US-056808.

PR 02-OCT-1997; US-061321.

PR 05-NOV-1997; US-064866.

PA (RIBO-) RIBOZYME PHARM INC.

PI Beaudry A, Beigelman L, Bellon L, Burgin A, Jarvis T,

PI Karpeisky A, Kisich K, Matulic-Adamic J, McSwiggen JA,

PI Parry T, Reynolds M, Svedler D, Thompson J, Workman CT;

PI WPI; 99-009494/01.

PT Identifying new catalytic nucleic acid that modulates selected  
 PT processes - especially ribozymes that cleave Raf RNA for treating  
 PT cancer, restenosis, and also new ribozymes and modified nucleoside  
 PT triphosphates used as antiviral agents and synthons  
 PS Claim 17; Page 171; 259pp; English.

CC A method has been developed for the identification of a nucleic acid  
 CC capable of modulating a process in a biological system. The method  
 CC comprises: (a) introducing into the system a random library of nucleic  
 CC acid catalysts (NAC) having a substrate binding domain (SBD), comprising  
 CC a random sequence, and a catalytic domain (CD); and (b) identifying NAC  
 CC in systems where modulation has occurred and/or determining the sequence  
 CC of at least part of the SBDs in such systems. Nucleic acid molecules with  
 CC endonuclease activity and catalytic activity, from the present invention,  
 CC are used to modulate gene expression in plant and mammalian cells and to  
 CC cleave target nucleic acid, particularly for treating systemic diseases  
 CC caused by specific RNA, e.g. cancer, inflammation, psoriasis, non-hepatic  
 CC ascites and infection. They may also be used to detect genetic drift and  
 CC mutations in diseased cells and to determine c-raf RNA. Specifically NACs  
 CC with RNA-cleaving activity that modulate expression of the Raf gene, are  
 CC used to treat cancer, restenosis, psoriasis or rheumatoid arthritis, or

CC generally any condition associated with the level of c-raf. Introduction  
 CC of sugar/phosphate modifications increases stability against nuclease  
 CC and activity. V90922 to V93877 represent NACs that can be used in the  
 CC method, specifically for modulating the expression of a Raf gene.  
 SQ Sequence 17 BP; 5 A; 3 C; 5 G; 4 U;

Query Match 59.0%; Score 11.8; DB 1; Length 17;  
 Best Local Similarity 86.7%; Pred. No. 1e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 ccgtacctgacttag 18  
 ||||| ||||| |||||  
 Db 17 CCGTACCTTACTGAG 3

## RESULT 7

T95409/c  
 ID T95409 standard; DNA; 20 BP.

AC T95409;

DT 10-MAR-1998 (first entry)

DE Forward primer for breast cancer susceptibility gene BRCA2 exon 6.

KW Human; breast cancer; susceptibility; gene; BRCA2; diagnosis;

KW screening; treatment; gene therapy; PCR primer; exon 6; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9722689-A1.

PD 26-JUN-1997.

PF 17-DEC-1996; U19598.

PR 29-APR-1996; US-639501.

PR 18-DEC-1995; US-573779.

PR 20-DEC-1995; US-575359.

PR 21-DEC-1995; US-576559.

PR 11-JAN-1996; US-585391.

PA (HSCR-) HSC RES & DEV LP.

PA (MYRI-) MYRIAD GENETICS I.

PA (UYPE-) UNIV PENNSYLVANIA.

PA (ENDO-) ENDO RECH INC.

PI Couch F, Kamb A, Rommens JM, Simard J, Tavtigian SV;

PI Weber BL;

DR WPI; 97-341680/31.

PT Human breast cancer susceptibility gene BRCA2 - useful for

PT diagnosing breast cancer and screening for compounds to treat breast

PT cancer

PS Example 3; Page 60; 189pp; English.

CC The present sequence is a primer for the human breast cancer

CC susceptibility gene BRCA2, which can be used to diagnose breast

CC cancer and screen for compounds to treat breast cancer. BRCA2 can

CC also be used in gene therapy to restore wild type BRCA2 gene

CC function to a cell, which has lost its or has altered (i.e. by

CC virtue of a mutation in BRCA2) BRCA2 gene function.

SQ Sequence 20 BP; 7 A; 1 C; 6 G; 6 T;

Query Match 59.0%; Score 11.8; DB 1; Length 20;  
 Best Local Similarity 86.7%; Pred. No. 1.1e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 ccgtacctgacttagc 19  
 ||||| ||||| |||||  
 Db 17 CATACCTGACTTATC 3

## RESULT 8

Q37759/c  
 ID Q37759 standard; DNA; 21 BP.

AC Q37759;

DT 29-JUN-1993 (first entry)

DE VMOI C-terminal primer (amino acids W123-W129).

KW VMOI; vitelline membrane outer I; primer; PCR; ss.

OS Synthetic.

PN J05023186-A.

PD 02-FEB-1993.

```

PF 19-JUL-1991; 179307.
PR 19-JUL-1991; JP-179307.
PA (TANP-) TANPAKU KOGAKU KENKYUSHO KK.
DR WPI; 93-07947/10.
PT DNA sequence coding VMOI - prep'd. e.g. by collecting fresh yolk
PT membrane, extracting with sodium chloride, dialysing and
PT subjecting to CM-Toyopearl 650M column chromatography, etc.
PS Disclosure; Fig 9; 28pp; Japanese.
CC The primers of Q37755-56 are used in the detection and isolation of
CC VMOI cDNA.
SQ Sequence 21 BP; 4 A; 9 C; 6 G; 2 T;

Query Match 59.0%; Score 11.8; DB 1; Length 21;
Best Local Similarity 86.7%; Pred. No. 1.1e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtaccgtgacttagcc 20
Db 20 GTACCGGCTTAGCC 6

RESULT 9
ID V70405 standard; DNA; 27 BP.
AC V70405;
DT 01-MAR-1999 (first entry)
DE Hepatitis delta virus anti-genome ribozyme primer.
KW HPIV-3; human parainfluenza virus type 3; vaccine; ribozyme;
KW PCR; primer; ss.
OS Synthetic.
OS Hepatitis delta virus.
PN W09850405-A1.
PD 12-NOV-1998.
PF 06-MAY-1998; U09270.
PR 07-MAY-1997; US-045805.
PA (CLEV-) CLEVELAND CLINIC FOUND.
PI Banerjee AK, Hoffman MA;
DR WPI; 98-610318/51.
PT New human parainfluenza virus (HPIV) clone, and - methods of clone
PT production and introduction of site-specific mutations
PS Example 1; 7pp; 54pp; English.
CC This primer encodes the 3' end of a hepatitis delta virus
CC antigenomic ribozyme and a BglII site. It was used with a second
CC primer (see V70404), which contains a sequence from the 3'
CC untranslated region of human parainfluenza virus type 3 (HPIV-3) L
CC mRNA, through to the trailer portion of the HPIV-3 genome (see
CC V70404), and which also primes synthesis of the ribozyme, in the
CC PCR amplification of the ribozyme. The product was used in the
CC construction of a full-length cDNA clone of HPIV-3. The invention
CC relates to a system for generating recombinant HPIV such as a clone
CC comprising a nucleotide sequence that encodes a full-length,
CC positive sense, anti-genome of HPIV. The invention also relates to
CC a method of preparing recombinant HPIV-3 having site-directed
CC mutations. Production of recombinant HPIV clones enables analysis
CC of HPIV transcription and replication in a cell-free system. A
CC system which permits production of recombinant HPIV that is
CC genetically engineered to contain site-specific mutations is useful
CC for identifying attenuating genotypes useful for developing live
CC vaccines.
SQ Sequence 27 BP; 4 A; 10 C; 6 G; 7 T;

Query Match 59.0%; Score 11.8; DB 1; Length 27;
Best Local Similarity 86.7%; Pred. No. 1.1e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtaccgtgacttagcc 20
Db 5 GTACCTCCCTTAGCC 19

RESULT 10

```

```

Q78691
ID Q78691 standard; DNA; 30 BP.
AC Q78691;
DT 24-JUN-1995 (first entry)
DE DNA probe for human IgA membrane anchor.
KW Probe; IgA membrane anchor; IgA epitope; monoclonal antibody;
KW therapeutic; ds.
OS Homo sapiens.
PN US5362643-A.
PD 08-NOV-1994.
PF 21-JUN-1989; 369479.
PR 21-JUN-1989; US-369479.
PR 22-DEC-1989; US-455080.
PR 16-SEP-1991; US-760765.
PR 20-JUL-1993; US-095068.
PA (TANO-) TANOX BIOSYSTEMS.
PI Chang TW;
DR WPI; 94-357359/44.
PT New antibodies specific for membrane bound IgA - and hybridomas
PT producing them, used to increase IgA prodn., partic. in patients
PT with infectious disease or allergy.
PS Disclosure; Page 9; 12pp; English.
CC The probe corresponds to a segment located in the human IgA
CC membrane anchor CH3 coding region of Ig allotype alpha-1 and
CC alpha-2, and is used to screen a human lung fibroblast line
CC (WI38) phage library containing alpha-1 or alpha-2 gene segments
CC by in situ hybridization. The human IgA membrane-anchoring
CC extracellular peptide segment is used to induce the formation of
CC monoclonal antibodies which modulate (increase) IgA synthesis.
CC The IgA can be used in the treatment of patients subject to
CC infectious diseases or suffering from allergy.
SQ Sequence 30 BP; 6 A; 9 C; 10 G; 5 T;

Query Match 59.0%; Score 11.8; DB 1; Length 30;
Best Local Similarity 86.7%; Pred. No. 1.1e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtaccgtgacttagcc 20
Db 8 GTACCTGACTTGGCC 22

RESULT 11
T10549
ID T10549 standard; DNA; 30 BP.
AC T10549;
DT 18-JUL-1996 (first entry)
DE Human IgA membrane anchoring extracellular peptide segment probe.
KW Exon; membrane anchoring extracellular peptide; human; immunoglobulin;
KW IgA; constant heavy region; cell surface; lung fibroblast cell line;
KW primer; PCR; amplification; probe; isoform; splicing site; antibody;
KW post-transcriptional processing; prophylaxis; infectious disease;
KW allergy; immunodeficiency disease; ds.
OS Synthetic.
PN US5484907-A.
PD 16-JAN-1996.
PF 21-JUN-1989; 369479.
PR 21-JUN-1989; US-369479.
PR 22-DEC-1989; US-455080.
PR 16-SEP-1991; US-760765.
PR 20-JUL-1993; US-095068.
PR 22-OCT-1993; US-140721.
PA (TANO-) TANOX BIOSYSTEMS INC.
PI Chang NT, Chang TW;
DR WPI; 96-087117/09.
PT Oligo:nucleotide(s) corresponding to human IgA segments - comprising
PT membrane anchoring extracellular peptide segments, used to develop
PT prods. for therapy and diagnosis
PS Example 1; Column 15; 12pp; English.
CC This is the nucleotide sequence of a probe used to isolate phage clones
CC contg. the sequences encoding the alpha-1 and alpha-2 isoforms of the
CC membrane anchoring peptide from a human IgA. The probe was used to screen

```

CC a human lung fibroblast line library in the phage FIX. The sequences  
 CC encoding the extracellular portion of the membrane anchoring peptide  
 CC (R88191) can be used to raise antibodies against the IgA membrane  
 CC extracellular peptide which can modulate IgA synthesis, esp. to increase  
 CC their prodn. The peptides and antibodies can be used to treat or in the  
 CC prophylaxis of infectious diseases, allergies and immunodeficiency  
 CC diseases.  
 SQ Sequence 30 BP; 6 A; 9 C; 10 G; 5 T;

Query Match 59.0%; Score 11.8; DB 1; Length 30;  
 Best Local Similarity 86.7%; Pred. No. 1.1e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtacctgacttagcc 20  
 ||||| ||||| |||||  
 DB 8 GTACCTGACTTGGGC 22

## RESULT 12

T86542  
 ID T86542 standard; DNA; 30 BP.  
 AC T86542;  
 DT 20-MAR-1998 (first entry)  
 DE Membrane extracellular peptide fragment of immunoglobulin probe.  
 KW Membrane bound; immunoglobulin A; anti-IgA antibody; immunogen;  
 KW B-cell leukemia; lymphoma; IgA-mediated nephropathy; diagnosis;  
 KW PCR; primer; probe; ss.  
 OS Homo sapiens.  
 PN US5690934-A.  
 PD 25-NOV-1997.  
 PF 20-MAR-1996; 619790.  
 PR 20-MAR-1996; US-619790.  
 PR 31-DEC-1987; US-140036.  
 PR 29-JUL-1988; US-226421.  
 PR 05-AUG-1988; US-229178.  
 PR 16-NOV-1988; US-272243.  
 PR 21-JUN-1989; US-369479.  
 PR 21-JUN-1989; US-369625.  
 PR 22-DEC-1989; US-455080.  
 PR 23-JAN-1990; US-468766.  
 PR 27-APR-1990; US-515604.  
 PR 29-OCT-1992; US-973321.  
 PR 09-JUL-1993; US-090527.  
 PR 20-JUL-1993; US-095068.  
 PR 14-OCT-1993; US-137253.  
 PR 22-OCT-1993; US-140721.  
 PR 11-JAN-1994; US-180145.  
 PR 26-MAY-1994; US-249558.  
 PA (TANO-) TANOX BIOSYSTEMS INC  
 PI Chang NT, Chang TW;  
 DR WPI: 98-017568/02.

PT Peptide fragments of human membrane-bound immunoglobulin A - for  
 PT generating anti-IgA antibodies, useful for treatment of B-cell  
 PT leukaemia(s) or lymphoma(s) or IgA-mediated nephropathy  
 PS Example 1: Columns 11-12: 10pp; English.  
 CC This sequence represents an oligonucleotide which corresponds to a  
 CC segment located in the CH3 coding region of immunoglobulin allotype  
 CC alpha1 and alpha2 and was synthesised and used as a probe to screen  
 CC phage clones containing oil or alpha2 gene segments. The library was  
 CC constructed using genomic DNA from human lung fibroblast line,  
 CC W138, packaged in phage FIX. The invention relates to a unique  
 CC extracellular peptide segment present on B cell-bound but not secreted  
 CC IgA. These extracellular peptide segments form, entirely or in part,  
 CC antigenic epitopes unique to membrane bound but not secreted IgA, and  
 CC thereby provide a unique epitope on the IgA-bearing B cells to which  
 CC membrane bound IgA is attached. These peptide segments can be used as  
 CC immunogens to generate antibodies which specifically target  
 CC membrane-bound IgA and IgA-bearing B cells.  
 SQ Sequence 30 BP; 6 A; 9 C; 10 G; 5 T;

Query Match 59.0%; Score 11.8; DB 1; Length 30;

Best Local Similarity 86.7%; Pred. No. 1.1e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 6 gtacctgacttagcc 20  
 ||||| ||||| |||||  
 DB 8 GTACCTGACTTGGGC 22

## RESULT 13

V21467  
 ID V21467 standard; DNA; 31 BP.  
 AC V21467;  
 DT 17-AUG-1998 (first entry)  
 DE Murine protein targeting to glycogen gene PCR primer.  
 KW Protein targeting to glycogen; pMG; pPIC; mouse; diabetes;  
 KW animal model; knockout mouse; PCR; primer; ss.  
 OS Synthetic.  
 OS Mus musculus.  
 PN W09808948-A1.  
 PD 05-MAR-1998.  
 PF 22-AUG-1997; U14142.  
 PR 12-AUG-1997; US-055243.  
 PR 30-AUG-1996; US-025107.  
 PA (WARN ) WARNER LAMBERT CO.  
 PI Brady MJ, Printen JA, Saltiel AR;  
 DR WPI: 98-230264/20.  
 PT Isolated protein targeting to glycogen genes - used to develop  
 PT products for increasing glycogen levels in cells and to develop  
 PT animal models for the study of e.g. diabetes  
 PS Disclosure: Page 30; 71pp; English.  
 CC PCR primers (V21466 and V21467), which contain BamHI and KpnI  
 CC sites, respectively, were used to amplify a 2.0 kb region 5' to the  
 CC coding region of the gene (see V21462) coding for murine protein  
 CC targeting to glycogen (PTG) protein (see W54164). PTG is a novel  
 CC glycogen binding subunit of PPIC that is expressed predominantly in  
 CC insulin-sensitive tissues and which mediates the hormonal control  
 CC of glycogen accumulation in intact cells. A PTG knockout vector  
 CC was constructed. A mouse having its PTK gene deleted from its  
 CC genome is claimed. Such knockout mice are useful as animal models  
 CC for increasing glycogen in cells, and are specifically useful as a  
 CC model of diabetes.  
 SQ Sequence 31 BP; 9 A; 9 C; 7 G; 6 T;

Query Match 59.0%; Score 11.8; DB 1; Length 31;  
 Best Local Similarity 86.7%; Pred. No. 1.1e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtacctgacttagcc 20  
 ||||| ||||| |||||  
 DB 5 GTACCTGACTGAGCC 19

## RESULT 14

N82409  
 ID N82409 standard; DNA; 14 BP.  
 AC N82409;  
 DT 19-NOV-1990 (first entry)  
 DE Probe (II) for detecting the atrial vasodilation-like peptide  
 DE encoding sequence.  
 KW Hypertension; atrial vasodilation-like peptide; AVD-like peptide;  
 KW probe; ss.  
 OS Homo sapiens.  
 PN J63107996-A.  
 PD 12-MAY-1988.  
 PF 18-SEP-1986; 220834.  
 PR 18-SEP-1986; JP-220834.  
 PA (MITU) Mitsubishi Chem Ind KK.  
 DR WPI: 88-170990/25.  
 PT Peptide for treating hypertension - contains aminoacid sequence based  
 PT on asparagine, proline, methionine, tyrosine, alanine, serine, leucine,  
 PT lysine, histidine, etc.  
 PS Disclosure; pp; Japanese.

CC Another probe very similar to this one is represented in N81663.  
CC See also N81661 and N81662.  
SQ Sequence 14 BP; 2 A; 5 C; 1 G; 5 T; 1 Others;

Query Match 58.0%; Score 11.6; DB 1; Length 14;  
Best Local Similarity 85.7%; Pred. No. 1.3e+03;  
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 tacctgacttagcc 20  
||||| ||||:|  
Db 1 TACCTGCTTADCC 14

RESULT 15  
Q82147/c  
ID Q82147 standard; DNA; 21 BP.  
AC Q82147;  
DT 01-SEP-1995 (first entry)  
DE Chromosome 11 (locus D11S1054) STS primer CSRL-2f10-tA.  
KW sequence sampled mapping; genomic analysis; complex genome mapping;  
KW cosmid library; chromosome 11; sequence tagged site; STS analysis; ss.  
OS Synthetic.  
PN WO9429486-A.  
PD 22-DEC-1994.  
PF 15-JUN-1994; U06810.  
PR 15-JUN-1993; US-078471.  
PR 07-SEP-1993; US-117952.  
PA (SALK ) SALK INST BIOLOGICAL STUDIES.  
PI EVANS GA, Smith MW;  
DR WPI; 95-036508/05.  
PT Sequencing complex genomes, present as fragments in a cosmid  
PT library - by sequencing end-specific nucleotides of each clone  
PT then correlating with spatial relationship of cosmid, esp. for  
PT mammalian chromosomes.  
PS Example 4; Page 68; 128pp; English.  
CC Sequences were determined from the ends of chromosome 11-specific  
CC cosmid by automated sequencing without intermediate subcloning.  
CC A sample of 371 DNA sequence fragments were determined and of  
CC these, 277 were suitable for STS primer prediction by computer  
CC analysis (using the "Primer" program available from E.Lander, MIT).  
CC The STSS and cosmids were mapped by in situ hybridisation, somatic  
CC cell hybrid analysis or both. Using this method, 370 STSS specific  
CC for human chromosome 11 were generated and most of them were  
CC regionally mapped. This procedure illustrates a novel method for  
CC sequencing complex genomes, designated "sequence sampled mapping".  
CC The sequence sampled mapping method is useful for the completion of  
CC high density sequence-based maps, and ultimately, for the complete  
CC sequencing of genomic DNA directly from cosmid clones.  
CC See Q82001-Q82706 for STS primers.  
SQ Sequence 21 BP; 9 A; 7 C; 4 G; 1 T;

Query Match 58.0%; Score 11.6; DB 1; Length 21;  
Best Local Similarity 77.8%; Pred. No. 1.4e+03;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttag 18  
||||| |||||  
Db 18 TTGCTGTTCTGTGTAG 1

Search completed: May 23, 2000, 11:21:33  
Job time: 6015 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 11:00:17 : Search time 1337.41 Seconds  
(without alignments)  
60.613 Million cell updates/sec

Title: US-08-945-805-2

Perfect score: 20

Sequence: 1 ttgcgtagctgacttagcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 23194

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*

- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_est45:\*
- 80: gb\_est46:\*
- 81: gb\_est47:\*
- 82: gb\_gss1:\*
- 83: gb\_gss2:\*
- 84: gb\_gss3:\*
- 85: gb\_gss4:\*
- 86: em\_gss1:\*
- 87: em\_gss2:\*
- 88: em\_gss3:\*
- 89: em\_gss4:\*
- 90: gb\_gss5:\*
- 91: gb\_gss6:\*
- 92: gb\_gss7:\*
- 93: gb\_gss8:\*
- 94: gb\_gss9:\*
- 95: em\_gss5:\*
- 96: em\_gss6:\*
- 97: em\_gss7:\*
- 98: em\_gss8:\*
- 99: em\_gss9:\*
- 100: em\_gss10:\*
- 101: em\_gss11:\*
- 102: gb\_gss10:\*
- 103: gb\_gss11:\*
- 104: em\_gss12:\*
- 105: gb\_gss12:\*
- 106: gb\_gss13:\*
- 107: gb\_gss14:\*
- 108: gb\_gss15:\*
- 109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
C 1	13.4	67.0	30	79	AW248139	AW248139 2819841.5
C 2	11	55.0	34	26	W70854	W70854 md391f05.r1
C 3	11	55.0	49	45	A1348359	A1348359 om48a10.s
C 4	11	55.0	49	45	A1348359	A1348359 qo20f09.x
C 5	10.6	53.0	31	41	A1021071	A1021071 ua99f05.r
C 6	10.4	52.0	31	39	AA883594	AA883594 al46d08.s
C 7	10.4	52.0	31	42	A1154164	A1154164 ud78h05.r
C 8	10.4	52.0	34	60	A1789638	A1789638 un19d11.r
C 9	10.4	52.0	37	34	AA466917	AA466917 vfl0g04.r
C 10	10.4	52.0	48	22	R59822	R59822 yhl1d05.r1
C 11	10.4	52.0	49	30	AA273875	AA273875 vb96b03.r
C 12	10.4	52.0	50	41	AU009984	AU009984 AU009984
C 13	10.2	51.0	34	43	A1188358	A1188358 qd08c03.x
C 14	10.2	51.0	40	44	A1302741	A1302741 qn17f10.x
C 15	10.2	51.0	46	25	N78054	N78054 yw71g05.r1
C 16	10	50.0	37	40	AA933761	AA933761 cm59g06.s
C 17	10	50.0	40	25	W12126	W12126 ma60g05.r1
C 18	10	50.0	43	62	A1877135	A1877135 vz76h09.r
C 19	10	50.0	50	24	H87539	H87539 yw14h12.s1
C 20	9.8	49.0	31	45	A1379850	A1379850 te29c07.x
C 21	9.8	49.0	37	45	A1358100	A1358100 qv96c06.x
C 22	9.8	49.0	37	79	AW245295	AW245295 2820036.3
C 23	9.8	49.0	43	34	AA532560	AA532560 nj44e10.s
C 24	9.6	48.0	25	39	AA905902	AA905902 ol88b06.s
C 25	9.6	48.0	31	40	AA912393	AA912393 ol99c01.s
C 26	9.6	48.0	35	79	AW246497	AW246497 2821639.3
C 27	9.6	48.0	37	28	AA064598	AA064598 zm13b12.s
C 28	9.6	48.0	37	42	A1119937	A1119937 uc22e11.r
C 29	9.6	48.0	49	42	A1095777	A1095777 qb30d06.x
C 30	9.4	47.0	31	21	T61693	T61693 yb86f04.r1
C 31	9.4	47.0	34	21	T72000	T72000 ye07e08.r1
C 32	9.4	47.0	34	23	H22671	H22671 ync7c01.r1
C 33	9.4	47.0	36	68	HSW008057	AL043207 Homo sapi
C 34	9.4	47.0	37	63	A1953517	A1953517 tg29f12.x
C 35	9.4	47.0	43	47	A1468217	A1468217 tg56f09.x
C 36	9.4	47.0	46	34	AA524270	AA524270 ng34d02.s
C 37	9.4	47.0	49	38	AA808591	AA808591 oa46h06.s
C 38	9.4	47.0	49	40	AA968655	AA968655 cq76b11.s
C 39	9.4	47.0	49	94	AQ074024	AQ074024 EP(3)3466
C 40	9.2	46.0	28	43	A1187132	A1187132 qe42f11.x
C 41	9.2	46.0	28	43	A1192544	A1192544 qe66b02.x
C 42	9.2	46.0	34	30	AA266770	AA266770 mz66d05.r
C 43	9.2	46.0	37	42	A1089361	A1089361 qb05h05.x
C 44	9.2	46.0	39	79	AW246101	AW246101 2821344.5
C 45	9.2	46.0	41	20	TI7556	TI7556 mps v120 Th

## ALIGNMENTS

RESULT 1	AW248139/c	AW248139	30 bp	mrna	EST	07-JAN-2000
LOCUS	2819841.5	prime NIH_MGC_7	Homo sapiens	cdna	clone IMAGE:2819841.5	
DEFINITION	mrna sequence.					
ACCESSION	AW248139					
VERSION	AW248139.1	GI:6591132				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 30)					
TITLE	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.					
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)					
COMMENT	Unpublished (1999)					
	On Dec 20, 1995 this sequence version replaced gi:1135382.					
	Other_ESTRs: 2819841.3prime					
	Contact: Robert Strausberg, Ph.D.					
	Tel: (301) 496-1550					
	Email: Robert_Strausberg@nih.gov					

Tissue Procurement: DCTD/DTF cDNA Library Preparation: Ling Hong/Rubium Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing project Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/brp/image/image.html Base Calling / Quality Scores: PHRED from University of Washington Genome Center Trimming: cross\_match from University of Washington Genome Center PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley Drosophila Genome Project. University of Washington Genome Center: http://www.genome.washington.edu Low Quality Sequence: 25 contiguous PHRED high quality bases following vector sequence. Very Low Quality Sequence: Trace file contained 30 contiguous distinct peaks following vector sequence.

Plate: LLCW2 row: K column: 10  
High quality sequence stop: 25.

FEATURES  
Location/Qualifiers  
1..30  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2819841"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 8 a 9 c 10 g 3 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 79; Length 30;  
Best Local Similarity 93.3%; Pred. No. 1.9e+03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgccgtacctgacct 16  
||||| |||||||  
Db 27 TGCCGGACCTGACTT 13

RESULT 2  
W70854  
LOCUS  
DEFINITION  
W70854 34 bp mrna EST 17-JUN-1996  
md91f05.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus CDNA  
clone IMAGE:385377.5, similar to SW:HCDH\_PIG P00348  
3-HYDROXYACYL-COA DEHYDROGENASE ;, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
W70854  
W70854.1 GI:1381006  
EST.  
house mouse.  
Mus musculus

REFERENCE  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenger, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Thelander, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HMMI Mouse EST Project  
Unpublished (1996)  
On May 18, 1995 this sequence version replaced gi:811187.  
Contact: Marra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:237209

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: mob.REGA+ET

High quality sequence stop: 1.

#### FEATURES

Location/Qualifiers  
1..34  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:385377"  
/clone\_lib="Soares mouse embryo NMME13.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGGTACCAATCTGAAGTGGGAGCGCCGCGAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 l]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

6 a 11 c 8 g 9 t

#### BASE COUNT

ORIGIN  
Query Match 55.0%; Score 11; DB 26; Length 34;  
Best Local Similarity 73.7%; Pred. No. 3.7e+04;  
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 2 tgcgcctactgactgacc 20  
||||| ||||| ||||| |||||  
Db 12 TGCCATACCTCATTGAGGC 30

#### RESULT 3

AA923178/c  
LOCUS  
DEFINITION  
AA923178 49 bp mRNA EST 23-JUN-1998  
IMAGE:1544250 3', similar to TR:O14598 O14598 TESTIS-SPECIFIC BASIC PROTEIN Y 1. ; contains MER22 repetitive element ;, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AA923178.1 GI:3070487  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished (1997)  
On Jan 14, 1998 this sequence version replaced gi:11797620.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Trace considered overall poor quality  
Insert Length: 606 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.

#### FEATURES

Location/Qualifiers  
1..49  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1544250"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

12 a 15 c 9 g 13 t

#### BASE COUNT

ORIGIN  
Query Match 55.0%; Score 11; DB 40; Length 49;  
Best Local Similarity 73.7%; Pred. No. 4e+04;  
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 2 tgcgcctactgactgacc 20  
||||| ||||| ||||| |||||  
Db 24 TGGCGTCCTGACTTCACC 6

#### RESULT 4

AI348359/c  
LOCUS  
DEFINITION  
AI348359 49 bp mRNA EST 01-FEB-1999  
Q020F09.x1 NCI-CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1909097 3', similar to SW:NMA\_HUMAN Q13145 PUTATIVE TRANSMEMBRANE PROTEIN NMA PRECURSOR. ;, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AI348359.1 GI:4085565  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Cloning by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
Insert Length: 696 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 1.

#### FEATURES

Location/Qualifiers  
1..49  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1909097"  
/clone\_lib="NCI-CGAP\_Lu5"  
/tissue\_type="carcinoid"





```

VERSION AA466917.1 GI:2193057
SOURCE EST.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 37)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
On Apr 26, 1996 this sequence version replaced gi:1283432.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..37
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:835350"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pSPORT; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dtt):
5'-CGGTGACCGTCGACCGTTTGTGTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
BASE COUNT 10 a 9 c 12 g 6 t
ORIGIN
Query Match 52.0%; Score 10.4; DB 34; Length 37;
Best Local Similarity 70.0%; Pred. No. 7.9e+04;
Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ttgcggtacctgactgacc 20
| | | | | | | | | |
Db 13 TCCGAAACCTGAGTGACC 32

RESULT 10
R59822 48 bp mRNA EST 24-MAY-1995
LOCUS yhl1d05.r1 Soares infant brain lntb Homo sapiens cDNA clone
DEFINITION IMAGE:43095 5' similar to gb|K01562|HUMCRH1 Human Ro RNA (rRNA);,;
rRNA sequence.
ACCESSION R59822.1 GI:830517
VERSION R59822
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 48)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On May 18, 1995 this sequence version replaced gi:811082.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: M13RPI
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..48
/organism="Homo sapiens"
/db_xref="GDB:415636"
/db_xref="taxon:9606"
/clone="IMAGE:43095"
/clone_lib="Soares infant brain lntb"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AACTGGAGATTCGCGCCGAGGAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 15 a 13 c 7 g 12 t 1 others
ORIGIN
Query Match 52.0%; Score 10.4; DB 22; Length 48;
Best Local Similarity 91.7%; Pred. No. 8.3e+04;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ccgtacctgact 15
| | | | | | | | | |
Db 6 CCGTACTGACT 17

RESULT 11
AA273875/c 49 bp mRNA EST 28-MAR-1997
LOCUS yd96b03.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone
DEFINITION IMAGE:764813 5' similar to SW:DPSD_CRIGR P27465 PHOSPHATIDYLSELINE
DECARBOXYLASE PROENZYME ;, mRNA sequence.
ACCESSION AA273875.1 GI:1912941
VERSION AA273875
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 49)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project

```

JOURNAL  
COMMENT

Unpublished (1996)  
On May 18, 1995 this sequence version replaced gi:811459.  
Contact: Marra M/Mouse EST Project  
WashU-RHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:465733  
Trace considered overall poor quality  
possible reversed clone: similarity on wrong strand  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 1.

FEATURES  
source

Location/Qualifiers  
1. .49  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:764813"  
/clone\_lib="Soares mouse lymph node NbMLN"  
/sex="male"  
/tissue\_type="lymph node"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5',

TGTTACCAATCTGAAGTGGGCGCGCATCTTTTTTTTTTTTTTTTTTTT  
3'); double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. RNA  
provided by Dr. Bertrand Jordan. Library constructed and  
normalized by Bento Soares and M.Fatima Bonaldo."

17 a 13 c 9 g 10 t

BASE COUNT  
ORIGIN

Query Match 52.0%; Score 10.4; DB 30; Length 49;  
Best Local Similarity 91.7%; Pred. No. 8.4e+04;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgcgcctacctg 12  
|||||  
Db 48 TTGCTGTACTCTG 37

RESULT 12  
AU009984

LOCUS AU009984 50 bp mRNA EST 31-JUL-1998  
DEFINITION AU009984 Schizosaccharomyces pombe late log phase cDNA  
Schizosaccharomyces pombe cDNA clone spc00765, mRNA sequence.

ACCESSION AU009984

VERSION AU009984.1 GI:3346664

SOURCE EST.

ORGANISM fission yeast.

Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
Schizosaccharomycetaceae; Schizosaccharomyces.

1 (bases 1 to 50)

Morimyo, M. and Mita, K.

Identification of expressed sequence tags of Schizosaccharomyces  
pombe

Unpublished (1998)

On Jan 17, 1998 this sequence version replaced gi:2043583.

Contact: Mitsuaki Morimyo

Genome Research Group

National Institute of Radiological Sciences

9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan

Email: morimyo@nirs.go.jp.

FEATURES  
source

Location/Qualifiers  
1. .50  
/organism="Schizosaccharomyces pombe"  
/strain="972"  
/db\_xref="taxon:4896"  
/clone="spc00765"  
/clone\_lib="Schizosaccharomyces pombe late log phase cDNA"  
/sex="h minus"  
/note="Vector: M13mp19; The cDNA library of  
Schizosaccharomyces pombe was prepared by cloning cDNA  
into the SmaI site of M13mp19 DNA and the direction of DNA  
sequences was not always from 5' to 3'. The cDNA data of  
Schizosaccharomyces pombe are available for searching on  
the World Wide Web. (URL, http://www.nirs.go.jp)"

9 a 12 c 11 g 18 t

BASE COUNT  
ORIGIN

Query Match 52.0%; Score 10.4; DB 41; Length 50;  
Best Local Similarity 70.0%; Pred. No. 8.4e+04;  
Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ttgcgcctacctgacctgacc 20

Db 13 TTGCTGTCTTCACGTTGCC 32

RESULT 13  
AII188358

LOCUS AII188358 34 bp mRNA EST 28-OCT-1998  
DEFINITION qd08c03.x1 Soares\_placenta\_8to9weeks\_2NBHP8to9W Homo sapiens cDNA  
clone IMAGE:1723108 3' similar to SW:UNR\_CAVPO P29174 UNR PROTEIN  
; mRNA sequence.

ACCESSION AII188358

VERSION AII188358.1 GI:3739567

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 34)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Insert Length: 954 Std Error: 0.00

Seg primer: -40UP from Gibco

High quality sequence stop: 1.

Location/Qualifiers

1. .34

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1723108"

/clone\_lib="Soares\_placenta\_8to9weeks\_2NBHP8to9W"

/dev\_stage="two placenta: one from 8 weeks and another

from 9 weeks post conception"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a

modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5',

TGTTACCAATCTGAAGTGGGCGCGCATCTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaldo."

11 a 11 c 4 g 8 t

## BASE COUNT

Search completed: May 23, 2000, 11:00:22

Qy	5
Db	9

Qy 5 cgtacctgacttagcc 20  
| | | | | | | | | |  
Db 9 CCTACCNCGCTGAGCC 24

Search completed: May 23, 2000, 11:00:22

Job time: 5904 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 11:19:44 ; Search time 68.55 Seconds  
(without alignments)  
37.924 Million cell updates/sec

Title: US-08-945-805-2  
Perfect score: 20  
Sequence: 1 ttgccgtacctgacttagcc 20  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 328916

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	79.0	39	2	US-08-484-192-23
2	12.6	63.0	24	3	US-08-723-726-2
3	12.6	63.0	33	1	US-08-345-756-4
4	12.6	63.0	33	2	US-08-625-198-4
5	12.4	62.0	32	3	US-08-529-190B-62
6	12	60.0	20	4	US-08-651-420A-7
7	12	60.0	20	3	US-08-628-001A-124
8	11.8	59.0	20	3	US-08-639-501-25
9	11.8	59.0	20	5	US-09-044-946-25
10	11.8	59.0	30	1	US-08-055-917-4
11	11.8	59.0	30	1	US-08-095-068-4
12	11.8	59.0	30	1	US-08-140-721A-4
13	11.8	59.0	30	2	US-08-619-790C-4
14	11.6	58.0	20	3	US-07-785-565A-4
15	11.6	58.0	21	3	US-08-117-952-147
16	11.6	58.0	23	6	PCT-US95-07085-3
17	11.6	58.0	40	1	PCT-US95-066-961-34
18	11.6	58.0	43	2	US-08-474-633A-94
19	11.6	58.0	43	2	US-08-474-633A-95
20	11.6	58.0	43	4	US-08-737-524B-18
21	11.6	58.0	43	4	US-08-737-524B-19
22	11.6	58.0	43	6	PCT-US93-02480-17
23	11.6	58.0	43	6	PCT-US93-02480-18
24	11.6	58.0	43	6	PCT-US95-08501-3
25	11.6	58.0	43	6	PCT-US95-08501-4
26	11.4	57.0	18	3	US-08-816-693A-21
27	11.4	57.0	22	5	US-08-755-587-53

28	11.4	57.0	50	4	US-08-793-170-18	Sequence 18, Appl
29	11.4	57.0	50	5	US-08-892-873-18	Sequence 18, Appl
c 30	11.2	56.0	16	4	US-08-599-455B-24	Sequence 24, Appl
31	11.2	56.0	19	1	US-07-841-662-17	Sequence 17, Appl
32	11.2	56.0	19	1	US-08-209-797-17	Sequence 17, Appl
33	11.2	56.0	19	2	US-08-669-685-17	Sequence 17, Appl
34	11.2	56.0	19	5	US-09-103-486-17	Sequence 17, Appl
35	11.2	56.0	19	6	PCT-US93-01557-17	Sequence 17, Appl
36	11.2	56.0	20	4	US-08-193-039B-16	Sequence 16, Appl
37	11.2	56.0	35	4	US-08-595-043A-55	Sequence 55, Appl
c 38	11.2	56.0	35	4	US-08-595-043A-56	Sequence 56, Appl
c 39	11	55.0	23	1	US-08-048-164A-21	Sequence 21, Appl
c 40	11	55.0	23	1	US-08-460-462-21	Sequence 21, Appl
c 41	11	55.0	23	2	US-08-460-457-21	Sequence 21, Appl
c 42	11	55.0	23	2	US-08-460-458-21	Sequence 21, Appl
c 43	11	55.0	23	3	US-08-460-455-21	Sequence 21, Appl
c 44	11	55.0	23	3	US-08-330-394A-21	Sequence 21, Appl
c 45	11	55.0	24	1	US-08-488-212A-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1  
US-08-484-192-23  
; Sequence 23, Application US/08484192  
; Patent No. 5756291  
; GENERAL INFORMATION:  
; APPLICANT: GRIFFIN, LINDA C.  
; APPLICANT: ALBRECHT, GLENN  
; APPLICANT: LATHAM, JOHN  
; APPLICANT: LEUNG, LAWRENCE  
; APPLICANT: VERMAAS, ERIC  
; APPLICANT: TOOLE, JOHN J.  
; TITLE OF INVENTION: APAMERS SPECIFIC FOR BIOMOLECULES AND  
; NUMBER OF SEQUENCES: 181  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,192  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/934,387  
; FILING DATE: 21-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GRACEY, NANCY J.  
; REGISTRATION NUMBER: 28,216  
; REFERENCE/DOCKET NUMBER: 246102002221  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-813-5600  
; TELEFAX: 415-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: misc\_difference  
; LOCATION: replace(20, "")

; OTHER INFORMATION: /note="This is a 60 nucleotide  
; OTHER INFORMATION: stretch of random sequences."  
US-08-484-192-23

Query Match 79.0%; Score 15.8; DB 2; Length 39;  
Best Local Similarity 85.0%; Pred. No. 6.4;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcgtacctgacttagcc 20  
||||| ||||| || |||

Db 14 TTGCGNACCTGAATCGCC 33

RESULT 2  
US-08-723-726-2  
; Sequence 2, Application US/08723726  
; Patent No. 5851521  
; GENERAL INFORMATION:  
; APPLICANT: BRANELLEC, Didier  
; APPLICANT: WALSH, Kenneth  
; APPLICANT: ISNER, Jeffrey M.  
; APPLICANT: DENEFFLE, Patrice  
; TITLE OF INVENTION: VIRAL VECTORS AND THEIR USE FOR TREATING  
; TITLE OF INVENTION: HYPERPROLIFERATIVE DISORDERS, IN PARTICULAR RESTENOSIS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: Mail Drop 3C43, P.O. Box 5093  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0997  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,726  
; FILING DATE: 30-SEP-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US96/04493  
; FILING DATE: 28-MAR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 95-04234  
; FILING DATE: 31-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitzky, Martin F.  
; REGISTRATION NUMBER: 29,699  
; REFERENCE/DOCKET NUMBER: ST95022A-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610)454-3816  
; TELEFAX: (610)454-3808  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 24 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Other nucleic acid  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-723-726-2

Query Match 63.0%; Score 12.6; DB 3; Length 24;  
Best Local Similarity 78.9%; Pred. No. 2.7e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ttgcgtacctgacttagcc 20  
|| ||| ||||| |||

Db 6 TGACGTGCCTGACTATGCC 24

RESULT 3  
US-08-345-756-4  
; Sequence 4, Application US/08345756  
; Patent No. 5633438  
; GENERAL INFORMATION:  
; APPLICANT: Baszczynski, Chris  
; APPLICANT: Barbour, Eric  
; APPLICANT: Hattori, Jiro  
; APPLICANT: Miki, Brian  
; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/345,756  
; FILING DATE: 22-NOV-1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, STEPHEN A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 33229/236/PIHI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 33 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-345-756-4

Query Match 63.0%; Score 12.6; DB 1; Length 33;  
Best Local Similarity 78.9%; Pred. No. 2.8e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgcgtacctgacttagcc 20  
||||| ||||| || |||

Db 12 TGCCTTACCGTCTCAGCC 30

RESULT 4  
US-08-625-198-4  
; Sequence 4, Application US/08625198  
; Patent No. 5756324  
; GENERAL INFORMATION:  
; APPLICANT: Baszczynski, Chris  
; APPLICANT: Barbour, Eric  
; APPLICANT: Hattori, Jiro  
; APPLICANT: Miki, Brian  
; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
US-08-625-198-4

COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/345,756  
FILING DATE: 22-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, STEPHEN A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 33229/236/PIHI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-625-198-4

Query Match 63.0%; Score 12.6; DB 2; Length 33;  
Best Local Similarity 78.9%; Pred. No. 2.8e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tgccgtacctgacttagcc 20  
||||| ||||| || |||||  
Db 12 TGCCCTTACCCGTCAGCC 30

RESULT 5  
US-08-529-190B-62/c  
; Sequence 62, Application US/08529190B  
; Patent No. 5833991  
; GENERAL INFORMATION:  
; APPLICANT: Masucci, Maria G.  
; TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES  
; TITLE OF INVENTION: CONFERRING INVISIBILITY TO THE IMMUNE SYSTEM  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: One Financial Center  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Wordperfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/529,190B  
; FILING DATE: 15-SEP-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE9501324-9  
; FILING DATE: 10-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US08/522,595  
; FILING DATE: 01-SEP-1995  
; ATTORNEY/AGENT INFORMATION:

; NAME: Williams, Ph.D., Kathleen A  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: 3255/53015  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-345-9100  
; TELEFAX: 617-345-9111  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
US-08-529-190B-62

Query Match 62.0%; Score 12.4; DB 3; Length 32;  
Best Local Similarity 92.9%; Pred. No. 3.6e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 cgtacctgacttag 18  
||||| ||||| |||||  
Db 28 CGTAACCTGACTTAG 15

RESULT 6  
US-08-651-420A-7/c  
; Sequence 7, Application US/08651420A  
; Patent No. 5969094  
; GENERAL INFORMATION:  
; APPLICANT: Compans, Richard W.  
; APPLICANT: Yao, Qizhi  
; TITLE OF INVENTION: Anti-paramyxovirus screening method and  
; TITLE OF INVENTION: vaccine  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,420A  
; FILING DATE: 22-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/135,285  
; FILING DATE: 12-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greenlee, Lorraine L  
; REGISTRATION NUMBER: 27,894  
; REFERENCE/DOCKET NUMBER: 35-93A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; HYPOTHETICAL: NO  
US-08-651-420A-7

Query Match 60.0%; Score 12; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gtacctgactta 17  
| | | | | | | | | |  
DB 12 GTACCTGACTTA 1

RESULT 7  
US-08-629-001A-124  
; Sequence 124, Application US/08629001A  
; Patent No. 5858661  
; GENERAL INFORMATION:  
; APPLICANT: Shilon, Yosef  
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS  
; TITLE OF INVENTION: GENOMIC ORGANIZATION  
; NUMBER OF SEQUENCES: 139  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kohn & Associates  
; STREET: 30500 No. 5858661thwestern Hwy.  
; CITY: Farmington Hills  
; STATE: Michigan  
; COUNTRY: US  
; ZIP: 48334  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/629,001A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kohn, Kenneth I.  
; REGISTRATION NUMBER: 30,955  
; REFERENCE/DOCKET NUMBER: 2290.00032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (810) 539-5050  
; TELEFAX: (810) 539-5055  
; INFORMATION FOR SEQ ID NO: 124:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-629-001A-124

Query Match 60.0%; Score 12; DB 3; Length 30;  
Best Local Similarity 75.0%; Pred. No. 5.7e+02;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgcctgactgacttagcc 20  
| | | | | | | | | |  
DB 8 TTCCCTTCTCTGGCTTAGCC 27

RESULT 8  
US-08-639-501-25/c  
; Sequence 25, Application US/08639501  
; Patent No. 5837492  
; GENERAL INFORMATION:  
; APPLICANT: Tavtligian, Sean V.  
; APPLICANT: Kamb, Alexander  
; APPLICANT: Simard, Jacques  
; APPLICANT: Couch, Fergus  
; APPLICANT: Rommens, Johanna  
; APPLICANT: Weber, Barbara  
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Venable, Baetjer, Howard & Civiletti  
; STREET: 1201 New York Avenue N.W., Suite 1001  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 22204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/639,501  
; FILING DATE: 29-APR-1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/585,391  
; FILING DATE: 11-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/576,559  
; FILING DATE: 21-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/575,359  
; FILING DATE: 20-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/573,779  
; FILING DATE: 18-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 24884-116802-04  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4810  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-639-501-25

Query Match 59.0%; Score 11.8; DB 3; Length 20;  
Best Local Similarity 86.7%; Pred. No. 7e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 cgtacctgacttagc 19  
| | | | | | | | | |  
DB 17 CATACCTGACTTATC 3

RESULT 9  
US-09-044-946-25/c  
; Sequence 25, Application US/09044946  
; Patent No. 6033857  
; GENERAL INFORMATION:  
; APPLICANT: Tavtligian, Sean V.  
; APPLICANT: Kamb, Alexander  
; APPLICANT: Simard, Jacques  
; APPLICANT: Couch, Fergus  
; APPLICANT: Rommens, Johanna  
; APPLICANT: Weber, Barbara  
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer  
; TITLE OF INVENTION: Susceptibility Gene  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti

STREET: 1201 New York Avenue N.W., Suite 1001  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 22204

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/044.946

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/639.501

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/576.559

FILING DATE: 21-DEC-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/575.359

FILING DATE: 20-DEC-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/573.779

FILING DATE: 18-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Ihnen, Jeffrey L.

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-116802-04

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

US-09-044-946-25

Query Match 59.0%; Score 11.8; DB 5; Length 20;  
Best Local Similarity 86.7%; Pred. NO. 7e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 cgtacctgacttagc 19

Db 17 CATACCTGACTTATC 3

## RESULT 10

US-08-055-917-4

Sequence 4, Application US/08055917

Patent No. 5310875

GENERAL INFORMATION:

APPLICANT: Chang, Tse Wen; Chang, Nancy T.

TITLE OF INVENTION: Peptides corresponding to membrane-bound IgA

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Tanox Biosystems, Inc.

STREET: 10301 Stella Link Rd.

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch

COMPUTER: IBM PS/2  
OPERATING SYSTEM: DOS 3.30  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/055.917  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/788.120  
FILING DATE: 11/4/1991  
APPLICATION NUMBER: 07/455.080  
FILING DATE: 12/22/1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirabel, Eric P.  
REGISTRATION NUMBER: 31,211  
REFERENCE/DOCKET NUMBER: TNX89-04CCC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 664-2288  
TELEFAX: (713) 664-8914  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: Double stranded  
TOPOLOGY: Linear  
US-08-055-917-4

Query Match 59.0%; Score 11.8; DB 1; Length 30;  
Best Local Similarity 86.7%; Pred. NO. 7.2e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 gtacctgacttagcc 20

Db 8 GTACCTGACTTGCGC 22

## RESULT 11

US-08-095-068-4

Sequence 4, Application US/08095068

Patent No. 5362643

GENERAL INFORMATION:

APPLICANT: Chang, Tse Wen; Chang, Nancy T.

TITLE OF INVENTION: Producing antibodies which bind to membrane-bound IgA using

TITLE OF INVENTION: an immunogen

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Tanox Biosystems, Inc.

STREET: 10301 Stella Link Rd.

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch

COMPUTER: IBM PS/2

OPERATING SYSTEM: DOS 3.30

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/095.068

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/760.765

FILING DATE: 9/16/1991

APPLICATION NUMBER: 07/455.080

FILING DATE: 12/22/1989

ATTORNEY/AGENT INFORMATION:

NAME: Mirabel, Eric P.

REGISTRATION NUMBER: 31,211

REFERENCE/DOCKET NUMBER: TNX89-04DEE

TELECOMMUNICATION INFORMATION:

TELEPHONE: (713) 664-2288

```
; TELEFAX: (713) 664-8914
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 30 nucleotides
;   TYPE: nucleic acid
;   STRANDEDNESS: Double stranded
;   TOPOLOGY: Linear
; US-08-095-068-4

Query Match          59.0%; Score 11.8; DB 1; Length 30;
Best Local Similarity 86.7%; Pred. No. 7.2e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtacctgacttagcc 20
Db 8 GTACCTGACTTGGC 22

RESULT 12
US-08-140-721A-4
; Sequence 4, Application US/08140721A
; Patent No. 5484907
; GENERAL INFORMATION:
; APPLICANT: Chang, Tse Wen; Chang, Nancy T.
; TITLE OF INVENTION: Nucleotides Coding for the Extracellular Membrane-Bound Segment
; TITLE OF INVENTION: IgA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tanox Biosystems, Inc.
; STREET: 10301 Stella Link Rd.
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS 3.30
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/140,721A
; FILING DATE: 7/20/1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirabel, Eric P.
; REGISTRATION NUMBER: 31,211
; REFERENCE/DOCKET NUMBER: TNX89-04FFF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 664-8914
; TELEFAX: (713) 664-8914
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 30 nucleotides
;   TYPE: nucleic acid
;   STRANDEDNESS: Double stranded
;   TOPOLOGY: Linear
; US-08-140-721A-4

Query Match          59.0%; Score 11.8; DB 1; Length 30;
Best Local Similarity 86.7%; Pred. No. 7.2e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtacctgacttagcc 20
Db 8 GTACCTGACTTGGC 22

RESULT 13
US-08-140-721A-4
; Sequence 4, Application US/08140721A
; Patent No. 5484907
; GENERAL INFORMATION:
; APPLICANT: Chang, Tse Wen; Chang, Nancy T.
; TITLE OF INVENTION: Nucleotides Coding for the Extracellular Membrane-Bound Segment
; TITLE OF INVENTION: IgA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tanox Biosystems, Inc.
; STREET: 10301 Stella Link Rd.
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS 3.30
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/140,721A
; FILING DATE: 7/20/1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirabel, Eric P.
; REGISTRATION NUMBER: 31,211
; REFERENCE/DOCKET NUMBER: TNX89-04FFF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 664-8914
; TELEFAX: (713) 664-8914
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 30 nucleotides
;   TYPE: nucleic acid
;   STRANDEDNESS: Double stranded
;   TOPOLOGY: Linear
; US-08-140-721A-4

Query Match          59.0%; Score 11.8; DB 1; Length 30;
Best Local Similarity 86.7%; Pred. No. 7.2e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtacctgacttagcc 20
Db 8 GTACCTGACTTGGC 22

RESULT 14
US-07-785-565A-4
; Sequence 4, Application US/07785565A
; Patent No. 5866129
; GENERAL INFORMATION:
; APPLICANT: Chang, Tse Wen; Chang, Nancy T.
; TITLE OF INVENTION: Treating Disease with a Peptide Corresponding to Membrane-B
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tanox Biosystems, Inc.
; STREET: 10301 Stella Link Rd.
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS 3.30
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/785,565A
; FILING DATE: 03/20/1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/249,558
; FILING DATE: 05/26/1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirabel, Eric P.
; REGISTRATION NUMBER: 31,211
; REFERENCE/DOCKET NUMBER: TNX89-04FGG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 664-2288
; TELEFAX: (713) 664-8914
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 30
;   TYPE: nucleic acid
;   STRANDEDNESS: Double
;   TOPOLOGY: Linear
; US-08-619-790C-4

Query Match          59.0%; Score 11.8; DB 2; Length 30;
Best Local Similarity 86.7%; Pred. No. 7.2e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 gtacctgacttagcc 20
Db 8 GTACCTGACTTGGC 22
```

; APPLICATION NUMBER: US/07/785,565A  
; FILING DATE: 19911104  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/455,080  
; FILING DATE: 12/22/1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mirabel, Eric P.  
; REGISTRATION NUMBER: 31,211  
; REFERENCE/DOCKET NUMBER: TNX89-04DDD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 664-2288  
; TELEFAX: (713) 664-8914  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 nucleotides  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: Double stranded  
; TOPOLOGY: Linear  
US-07-785-565A-4

Query Match 59.0%; Score 11.8; DB 3; Length 30;  
Best Local Similarity 86.7%; Pred. No. 7.2e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 gtacctgacttagcc 20  
|||||  
Db 8 GTACCTGACTTGGGC 22

RESULT 15  
US-08-117-952-147/c  
; Sequence 147, Application US/08117952  
; Patent No. 5851760  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; APPLICANT: Smith, Michael W.  
; TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE  
; TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES  
; NUMBER OF SEQUENCES: 797  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/117,952  
; FILING DATE: 07-SEP-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/078,471  
; FILING DATE: 15-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: P41 9423  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4737  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 147:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

; MOLECULE TYPE: Oligonucleotide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-117-952-147  
  
Query Match 58.0%; Score 11.6; DB 3; Length 21;  
Best Local Similarity 77.8%; Pred. No. 8.9e+02;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 ttgccgtacctgacttag 18  
|||||  
Db 18 TTGCTGTTCCTGTGTTAG 1  
  
Search completed: May 23, 2000, 11:19:47  
Job time: 5944 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 10:36:32 ; Search time 236.64 Seconds  
(without alignments)  
-82.217 Million cell updates/sec

Title: US-08-945-805-3  
Perfect score: 20  
Sequence: 1 ggaacttcctcctaaaggagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 285916

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_bal:\*
- 2: gb\_ba2:\*
- 3: gb\_ov:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_ph:\*
- 8: gb\_ph:\*
- 9: gb\_ph:\*
- 10: gb\_ph:\*
- 11: gb\_ph:\*
- 12: gb\_ph:\*
- 13: gb\_ph:\*
- 14: gb\_ph:\*
- 15: gb\_ph:\*
- 16: gb\_ph:\*
- 17: gb\_ph:\*
- 18: gb\_ph:\*
- 19: gb\_ph:\*
- 20: gb\_ph:\*
- 21: gb\_ph:\*
- 22: gb\_ph:\*
- 23: gb\_ph:\*
- 24: gb\_ph:\*
- 25: gb\_ph:\*
- 26: gb\_ph:\*
- 27: gb\_ph:\*
- 28: gb\_ph:\*
- 29: gb\_ph:\*
- 30: gb\_ph:\*
- 31: gb\_ph:\*
- 32: gb\_ph:\*
- 33: gb\_ph:\*
- 34: gb\_ph:\*
- 35: gb\_ph:\*
- 36: gb\_ph:\*
- 37: gb\_ph:\*
- 38: gb\_ph:\*
- 39: gb\_ph:\*
- 40: gb\_ph:\*
- 41: gb\_ph:\*
- 42: gb\_ph:\*
- 43: gb\_ph:\*
- 44: gb\_ph:\*

- 45: gb\_htg7:\*
- 46: em\_htg1:\*
- 47: em\_htg2:\*
- 48: em\_htg3:\*
- 49: em\_hum5:\*
- 50: gb\_pl3:\*
- 51: gb\_pr5:\*
- 52: gb\_htg8:\*
- 53: gb\_htg9:\*
- 54: gb\_htg10:\*
- 55: gb\_htg11:\*
- 56: gb\_htg12:\*
- 57: gb\_htg13:\*
- 58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	13.6	68.0	42	5	A36505	A36505 Sequence 46
2	13.2	66.0	45	5	AR032518	AR032518 Sequence
3	13.2	66.0	45	5	AR032730	AR032730 Sequence
4	13.2	66.0	45	5	I29258	I29258 Sequence 13
5	13.2	66.0	45	5	I29470	I29470 Sequence 34
6	13.2	66.0	45	5	I90932	I90932 Sequence 13
7	13.2	66.0	45	5	I91144	I91144 Sequence 34
8	13.2	66.0	45	5	AR032731	AR032731 Sequence
9	13.2	66.0	50	5	I29471	I29471 Sequence 34
10	13.2	66.0	50	5	I91145	I91145 Sequence 34
11	12.2	61.0	22	5	AR058722	AR058722 Sequence
12	12.2	61.0	23	5	I42208	I42208 Sequence 21
C 13	12.2	61.0	25	5	AR019488	AR019488 Sequence
C 14	12.2	61.0	25	5	E03560	E03560 Synthetic D
15	12.2	61.0	35	5	I93411	I93411 Sequence 9
16	12.2	61.0	35	5	I95038	I95038 Sequence 9
17	12.2	61.0	44	5	AR029775	AR029775 Sequence
C 18	12.2	61.0	48	5	A87403	A87403 Sequence 2
19	12.2	61.0	49	5	AR029766	AR029766 Sequence
20	12.2	61.0	49	5	AR058712	AR058712 Sequence
21	12.2	61.0	50	5	AR029769	AR029769 Sequence
C 22	12.2	60.0	25	11	HUMMOFRA	L32784 Homo sapien
23	12.2	60.0	29	5	I26549	I26549 Sequence 24
24	12.2	60.0	33	16	RAV2G2	K02863 Rabies viru
25	12.2	60.0	33	16	RAV2G2	K02867 Rabies viru
C 26	11.8	59.0	44	9	HUMTCVD1EV	L32497 Human (clon
C 27	11.8	59.0	45	10	HSU12318	U12318 Human T cel
C 28	11.8	59.0	45	10	HSU14047	U14047 Human T cel
C 29	11.8	59.0	46	5	A84816	A84816 Sequence 11
C 30	11.6	58.0	23	5	E12634	E12634 DNA oligome
31	11.6	58.0	24	5	A46132	A46132 Sequence 27
32	11.6	58.0	25	5	A50845	A50845 Sequence 18
33	11.6	58.0	30	5	I07195	I07195 Sequence 18
34	11.6	58.0	30	5	I30361	I30361 Sequence 15
35	11.6	58.0	33	16	RAV1G2	K02859 Rabies viru
36	11.6	58.0	42	5	AR000136	AR000136 Sequence
37	11.6	58.0	42	5	I66250	I66250 Sequence 2
38	11.6	58.0	45	5	AR032687	AR032687 Sequence
C 39	11.6	58.0	45	5	I19730	I19730 Sequence 3
40	11.6	58.0	45	5	I29427	I29427 Sequence 29
41	11.6	58.0	45	5	I91101	I91101 Sequence 29
42	11.6	58.0	46	5	E05194	E05194 Promoter . 9
43	11.6	58.0	46	5	E05581	E05581 Promoter . 9
44	11.4	57.0	20	5	AR016236	AR016236 Sequence
45	11.4	57.0	23	5	I12712	I12712 Sequence 10

ALIGNMENTS

RESULT 1  
A36505/c A36505 42 bp DNA PAT 05-MAR-1997  
LOCUS  
DEFINITION Sequence 46 from Patent W09323549.  
ACCESSION A36505  
VERSION A36505.1 GI:2293816  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 42)  
AUTHORS Aurias,A., Delattre,O., Desmaze,C., Melot,T., Peter,M.,  
Plougastel,B., Thomas,G. and Zucman,J.  
TITLE NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN  
RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE  
DEVELOPMENT OF CANCEROUS TUMORS  
JOURNAL CENTRE NAT RECH SCIENT (FR)  
Patent: WO 9323549-A 46 25-NOV-1993;  
COMMENT Other publication FR 2691475 931126  
Other publication JP 8500964T 960206.  
FEATURES  
source 1..42  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606" 14 t  
BASE COUNT 3 a 15 c 10 g 14 t  
ORIGIN  
  
Query Match 68.0%; Score 13.6; DB 5; Length 42;  
Best Local Similarity 80.0%; Pred. No. 4.1e+03;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 ggaacttcctcctaaaggagg 20  
||| ||||| |||||  
Db 39 GGAGACTCCCTAAAGTGAGG 20  
  
RESULT 2  
AR032518  
LOCUS  
DEFINITION Sequence 130 from patent US 5869241.  
ACCESSION AR032518  
VERSION AR032518.1 GI:5948123  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding  
molecule  
JOURNAL Patent: US 5869241-A 130 09-FEB-1999;  
FEATURES  
source 1..45  
Location/Qualifiers  
/organism="unknown" 9 t  
BASE COUNT 14 a 9 c 13 g 9 t  
ORIGIN  
  
Query Match 66.0%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 ggaacttcctcctaaaggagg 18  
||| ||||| |||||  
Db 12 GGAGCTTACATAAAGGGA 29  
  
RESULT 3  
AR032730  
LOCUS  
DEFINITION Sequence 342 from patent US 5869241.  
ACCESSION AR032730.1 GI:5948335  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding  
molecule  
JOURNAL Patent: US 5869241-A 342 09-FEB-1999;  
FEATURES  
source 1..45  
Location/Qualifiers  
/organism="unknown" 9 t  
BASE COUNT 14 a 9 c 13 g 9 t  
ORIGIN  
  
Query Match 66.0%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 ggaacttcctcctaaaggagg 18  
||| ||||| |||||  
Db 12 GGAGCTTACATAAAGGGA 29  
  
RESULT 5  
I29470  
LOCUS  
DEFINITION Sequence 342 from patent US 5578444.  
ACCESSION I29470  
VERSION I29470.1 GI:1820261  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Sequence-directed DNA-binding molecules compositions and methods  
JOURNAL Patent: US 5578444-A 342 26-NOV-1996;  
FEATURES  
source 1..45  
Location/Qualifiers  
/organism="unknown" 9 t  
BASE COUNT 14 a 9 c 13 g 9 t  
ORIGIN  
  
Query Match 66.0%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 ggaacttcctcctaaaggagg 18  
||| ||||| |||||  
Db 12 GGAGCTTACATAAAGGGA 29  
  
RESULT 5  
I29470  
LOCUS  
DEFINITION Sequence 342 from patent US 5578444.  
ACCESSION I29470  
VERSION I29470.1 GI:1820261  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Sequence-directed DNA-binding molecules compositions and methods  
JOURNAL Patent: US 5578444-A 342 26-NOV-1996;  
FEATURES  
source 1..45  
Location/Qualifiers  
/organism="unknown" 9 t  
BASE COUNT 14 a 9 c 13 g 9 t  
ORIGIN

DEFINITION Sequence 342 from patent US 5869241.  
ACCESSION AR032730  
VERSION AR032730.1 GI:5948335  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding  
molecule  
JOURNAL Patent: US 5869241-A 342 09-FEB-1999;  
FEATURES  
source 1..45  
Location/Qualifiers  
/organism="unknown" 9 t  
BASE COUNT 14 a 9 c 13 g 9 t  
ORIGIN  
  
Query Match 66.0%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 ggaacttcctcctaaaggagg 18  
||| ||||| |||||  
Db 12 GGAGCTTACATAAAGGGA 29  
  
RESULT 4  
I29258  
LOCUS  
DEFINITION Sequence 130 from patent US 5578444.  
ACCESSION I29258  
VERSION I29258.1 GI:1820049  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Sequence-directed DNA-binding molecules compositions and methods  
JOURNAL Patent: US 5578444-A 130 26-NOV-1996;  
FEATURES  
source 1..45  
Location/Qualifiers  
/organism="unknown" 9 t  
BASE COUNT 14 a 9 c 13 g 9 t  
ORIGIN  
  
Query Match 66.0%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 ggaacttcctcctaaaggagg 18  
||| ||||| |||||  
Db 12 GGAGCTTACATAAAGGGA 29  
  
RESULT 5  
I29470  
LOCUS  
DEFINITION Sequence 342 from patent US 5578444.  
ACCESSION I29470  
VERSION I29470.1 GI:1820261  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 45)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Sequence-directed DNA-binding molecules compositions and methods  
JOURNAL Patent: US 5578444-A 342 26-NOV-1996;  
FEATURES  
source 1..45  
Location/Qualifiers  
/organism="unknown" 9 t  
BASE COUNT 14 a 9 c 13 g 9 t  
ORIGIN

```
source 1. .45
/organism="unknown"
BASE COUNT 14 a 9 c 13 g 9 t
ORIGIN

Query Match 66.0%; Score 13.2; DB 5; Length 45;
Best Local Similarity 83.3%; Pred. No. 6.9e+03;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
||||| | | | | | | | |
Db 12 GGAGCTTACATAAGGGA 29

RESULT 6
I90932
LOCUS I90932 45 bp DNA PAT 01-DEC-1998
DEFINITION Sequence 130 from patent US 5726014.
ACCESSION I90932
VERSION I90932.1 GI:3935402
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 45)
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.
TITLE Screening assay for the detection of DNA-binding molecules
JOURNAL Patent: US 5726014-A 130 10-MAR-1998;
FEATURES Location/Qualifiers
source 1. .45
BASE COUNT 14 a 9 c 13 g 9 t
ORIGIN

Query Match 66.0%; Score 13.2; DB 5; Length 45;
Best Local Similarity 83.3%; Pred. No. 6.9e+03;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
||||| | | | | | | | |
Db 12 GGAGCTTACATAAGGGA 29

RESULT 7
I91144
LOCUS I91144 45 bp DNA PAT 01-DEC-1998
DEFINITION Sequence 342 from patent US 5726014.
ACCESSION I91144
VERSION I91144.1 GI:3935614
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 45)
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.
TITLE Screening assay for the detection of DNA-binding molecules
JOURNAL Patent: US 5726014-A 342 10-MAR-1998;
FEATURES Location/Qualifiers
source 1. .45
BASE COUNT 14 a 9 c 13 g 9 t
ORIGIN

Query Match 66.0%; Score 13.2; DB 5; Length 45;
Best Local Similarity 83.3%; Pred. No. 6.9e+03;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
||||| | | | | | | | |
Db 12 GGAGCTTACATAAGGGA 29
```

```
RESULT 8
AR032731
LOCUS AR032731 50 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 343 from patent US 5869241.
ACCESSION AR032731
VERSION AR032731.1 GI:5948336
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 50)
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.
TITLE Method of determining DNA sequence preference of a DNA-binding molecule
JOURNAL Patent: US 5869241-A 343 09-FEB-1999;
FEATURES Location/Qualifiers
source 1. .50
BASE COUNT 14 a 12 c 14 g 10 t
ORIGIN
```

```
Query Match 66.0%; Score 13.2; DB 5; Length 50;
Best Local Similarity 83.3%; Pred. No. 6.9e+03;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
||||| | | | | | | | |
Db 18 GGAGCTTACATAAGGGA 35
```

```
RESULT 9
I29471
LOCUS I29471 50 bp DNA PAT 06-FEB-1997
DEFINITION Sequence 343 from patent US 5578444.
ACCESSION I29471
VERSION I29471.1 GI:1820262
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 50)
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.
TITLE Sequence-directed DNA-binding molecules compositions and methods
JOURNAL Patent: US 5578444-A 343 26-NOV-1996;
FEATURES Location/Qualifiers
source 1. .50
BASE COUNT 14 a 12 c 14 g 10 t
ORIGIN
```

```
Query Match 66.0%; Score 13.2; DB 5; Length 50;
Best Local Similarity 83.3%; Pred. No. 6.9e+03;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
||||| | | | | | | | |
Db 18 GGAGCTTACATAAGGGA 35
```

```
RESULT 10
I91145
LOCUS I91145 50 bp DNA PAT 01-DEC-1998
DEFINITION Sequence 343 from patent US 5726014.
ACCESSION I91145
VERSION I91145.1 GI:3935615
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 50)
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.
TITLE Sequence-directed DNA-binding molecules compositions and methods
JOURNAL Patent: US 5726014-A 343 26-NOV-1996;
FEATURES Location/Qualifiers
source 1. .50
BASE COUNT 14 a 12 c 14 g 10 t
ORIGIN
```

REFERENCE 1 (bases 1 to 50)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.  
TITLE Screening assay for the detection of DNA-binding molecules  
JOURNAL Patent: US 5726014-A 343 10-MAR-1998;  
FEATURES Location/Qualifiers  
source 1..50  
BASE COUNT 14 a 12 c 14 g 10 t  
ORIGIN

Query Match 66.0%; Score 13.2; DB 5; Length 50;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaactccctaaaggga 18  
| | | | | | | | | | | | | | | | | | | | | |  
Db 18 GGAGCTTACATAAGGGA 35

RESULT 11  
AR058722  
LOCUS AR058722 22 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 299 from patent US 5837832.  
ACCESSION AR058722  
VERSION AR058722.1 GI:5984299  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Chee,M., Cronin,M.T., Fodor,S.P.A., Huang,X.X., Hubbell,E.A.,  
Lipshutz,R.J., Lohban,P.E., Morris,M.S. and Sheldon,E.L.  
TITLE Arrays of nucleic acid probes on biological chips  
JOURNAL Patent: US 5837832-A 299 17-NOV-1998;  
FEATURES Location/Qualifiers  
source 1..22  
BASE COUNT 9 a 5 c 4 g 4 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 22;  
Best Local Similarity 82.4%; Pred. No. 2.6e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aacttcctaaaggag 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 6 ACCCTCACTAAGGGAG 22

RESULT 12  
I42208  
LOCUS I42208 23 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 21 from patent US 5629153.  
ACCESSION I42208  
VERSION I42208.1 GI:2467703  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 23)  
AUTHORS Urdea,M.S.  
TITLE Use of DNA-dependent RNA polymerase transcripts as reporter  
molecules for signal amplification in nucleic acid hybridization  
assays  
JOURNAL Patent: US 5629153-A 21 13-MAY-1997;  
FEATURES Location/Qualifiers  
source 1..23  
BASE COUNT 9 a 5 c 4 g 5 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 23;  
Best Local Similarity 82.4%; Pred. No. 2.6e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aacttcctaaaggag 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 6 ACCCTCACTAAGGGAG 22

RESULT 13  
AR019488  
LOCUS AR019488 25 bp DNA PAT 05-DEC-1998  
DEFINITION Sequence 5 from patent US 5783442.  
ACCESSION AR019488  
VERSION AR019488.1 GI:3974602  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 25)  
AUTHORS Kato,S., Aoki,T. and Umezawa,Y.  
TITLE Cloning vector plasmid, vector-primer derived therefrom and  
preparation method of cDNA bank using the same  
JOURNAL Patent: US 5783442-A 5 21-JUL-1998;  
FEATURES Location/Qualifiers  
source 1..25  
BASE COUNT 4 a 4 c 7 g 10 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 25;  
Best Local Similarity 82.4%; Pred. No. 2.6e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aacttcctaaaggag 19  
| | | | | | | | | | | | | | | | | | | | | |  
Db 20 ACCCTCACTAAGGGAG 4

RESULT 14  
E03560/c  
LOCUS E03560 25 bp DNA PAT 29-SEP-1997  
DEFINITION Synthetic DNA oligomer.  
ACCESSION E03560  
VERSION E03560.1 GI:2171776  
KEYWORDS JP 1992117292-A/3.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 25)  
AUTHORS Kato,M., Aoki,T. and Umezawa,Y.  
TITLE CLONING VECTOR PLASMID, VECTOR PRIMER DERIVED FROM THE SAME PLASMID  
AND PREPARATION OF cDNA BANK USING THE SAME PRIMER  
JOURNAL Patent: JP 1992117292-A 3 17-APR-1992;  
COMMENT SAGAMI CHEM RES CENTER  
OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1992117292-A/3  
PD 17-APR-1992  
PF 07-SEP-1990 JP 1990238332  
PR 31-OCT-1989 JP 89P 283674  
PI KATO MASASHI, AOKI TAKASHI, UMEZAWA YURI  
PC C12N15/85,C12N15/70//C12N15/10;  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FEATURES Location/Qualifiers  
source 1..25  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 4 a 4 c 7 g 10 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 25;  
Best Local Similarity 82.4%; Pred. NO. 2.6e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aacttcctctaaaggag 19  
| | | | | | | | | |  
Db 20 ACCCTCACTAAGGGAG 4

RESULT 15  
193411  
LOCUS 193411 35 bp DNA PAT 01-DEC-1998  
DEFINITION Sequence 9 from patent US 5731144.  
ACCESSION 193411  
VERSION 193411.1 GI:3937881  
KEYWORDS .  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 35)  
AUTHORS Toothman,P.J., Ringquist,S. and Gold,L.  
TITLE High affinity TGF.beta. nucleic acid ligands  
JOURNAL Patent: US 5731144-A 9 24-MAR-1998;  
FEATURES Location/Qualifiers  
source 1..35  
/organism="unknown"  
BASE COUNT 12 a 8 c 8 g 7 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 35;  
Best Local Similarity 82.4%; Pred. NO. 2.6e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aacttcctctaaaggag 19  
| | | | | | | | | |  
Db 3 ACCCTCACTAAGGGAG 19

Search completed: May 23, 2000, 10:36:34  
Job time: 5935 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 11:21:33 ; Search time 96.59 Seconds  
(without alignments)  
51.805 Million cell updates/sec

Title: US-08-945-805-3

Perfect score: 20

Sequence: 1 ggaacttcctaaaggagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 391736

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	14.2	71.0	44	1 V80370	Heteropolymer olig
C 2	13.6	68.0	45	1 Q43300	Sequence of PCR pr
C 3	13.4	67.0	23	1 T73799	Multiplex short-PC
4	13.4	67.0	31	1 X39216	Human genomic DNA
5	13.4	67.0	31	1 X39369	Human genomic DNA
6	13.2	66.0	45	1 Q69380	Human fibrinogen g
7	13.2	66.0	45	1 Q69592	Human gene for fib
8	13.2	66.0	45	1 T64054	Human fibrinogen g
9	13.2	66.0	45	1 T63842	Human fibrinogen g
10	13.2	66.0	45	1 X17342	Test sequence from
11	13.2	66.0	45	1 X17130	Test sequence from
12	13.2	66.0	46	1 X36364	Primer for human T
13	13.2	66.0	48	1 X36360	Primer for human T
14	13.2	66.0	48	1 X32520	PCR primer for in
15	13.2	66.0	48	1 X52518	Human gene for fib
16	13.2	66.0	50	1 Q69593	Human fibrinogen g
17	13.2	66.0	50	1 T64055	Human gene for fib
18	13.2	66.0	50	1 X17343	Test sequence from
C 19	12.8	64.0	47	1 Q62707	Primer for amplifi
C 20	12.6	63.0	29	1 V62759	Human secreted pro
C 21	12.6	63.0	41	1 V25473	T7 promoter sequen
C 22	12.6	63.0	44	1 Q25657	CaMV35S minimal pr
C 23	12.6	63.0	47	1 T03932	Factor-V NASBA pri
C 24	12.6	63.0	47	1 T03932	3SR primer 325. Is
C 25	12.6	63.0	47	1 T03636	Primer BB325 for b
26	12.4	62.0	39	1 T92311	Wild-type CAT gene
27	12.4	62.0	39	1 V06784	Sequence of wild t
28	12.4	62.0	39	1 T86422	Wild type CAT gene
29	12.2	61.0	22	1 V39121	T3 promoter sequen
30	12.2	61.0	27	1 Q11892	ABS 62 T3 promoter
31	12.2	61.0	27	1 Q11889	ABS 63 T3 promoter
32	12.2	61.0	29	1 T76766	T3 RNA polymerase
33	12.2	61.0	31	1 V60718	Primer #1 for yeast
C 34	12.2	61.0	35	1 Q13184	T3 promoter. Dual

35 12.2 61.0 35 1 T65411 Transforming growt  
36 12.2 61.0 46 1 Q57604 M. tuberculosis rR  
c 37 12.2 61.0 48 1 V07636 Human leukocyte an  
38 12.2 61.0 48 1 X33612 EGF-like/FGF-8 hom  
39 12.2 61.0 48 1 X33602 EGF-like/FGF-8 hom  
40 12.2 61.0 48 1 X33604 PCR primer for in  
41 12.2 61.0 48 1 X52522 PCR primer for in  
42 12.2 61.0 48 1 X52530 PCR primer for in  
43 12.2 61.0 48 1 X52512 PCR primer for in  
44 12.2 61.0 48 1 X52514 PCR primer for in  
45 12.2 61.0 49 1 Q88404 p53 exon 5 T3 ampli

#### ALIGNMENTS

RESULT 1

V80370/c  
ID V80370 standard; DNA; 44 BP.  
AC V80370;  
DT 03-MAR-1999 (first entry)  
DE Heteropolymer oligo 2 used for detecting anti-SV40 analyte.  
KW Immunoassay; analyte; ligand; binding; immune complex; polymerase;  
KW polynucleotide assay reagent; PAR; cleavage; clinical; assay;  
KW detection; human immune deficiency virus; HIV; anti-SV40; ss.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT modified\_base 1  
FT /\*tag= a  
FT /note= "aminoalkylation with 1,2-diaminoethane"

PN US5849478-A.  
PD 15-DEC-1998.

PP 24-DEC-1992; 996793

PR 24-DEC-1992; US-996793.

PR 14-AUG-1986; US-897142.

PR 17-NOV-1988; US-272648.

PR 01-APR-1990; US-508259.

PA (CASH/) CASHMAN D P.

PI Cashman DP.

DR WPI; 99-069714/06.

PT Immunoassay using a polynucleotide linked to antigen ligand close to its initiation site - and able to be copied, in presence of polymerase, only if the initiation site is not blocked by bound analyte

PS Example 10; Column 27; 23pp; English.

CC The invention provides immunoassay methods for detecting an analyte in a liquid sample. The analyte is an antigen (Ag)-specific antibody (Ab) or an Ag that competes with an antigen ligand for binding to Ab. The method comprises first treating the sample with a binding reagent, including a polynucleotide assay reagent (PAR) comprising (i) antigen ligand and (ii) attached polynucleotide which has an initiation region adjacent to the ligand. The ligand can bind to an Ab to form an immune complex in which the initiation region is blocked and Ab is either the analyte or a component of binding reagent (in which case the analyte is an Ag that competes with the ligand for binding). After this step, PAR is treated with polymerase and nucleotide triphosphates so that the polynucleotide is copied, with production of cleavage products (containing a (pyro) phosphate group) only if the initiation region is not blocked. The reaction mixture is analysed for cleavage products, the amount of which is related to the amount of analyte in the sample. The method is suitable for a wide range of clinical assays, e.g detection of Ab specific for human immune deficiency virus (HIV) or other antigens. The method provides rapid, sensitive and precise detection of analytes, optionally several different analytes simultaneously. The present sequence represents a heteropolymer oligonucleotide used in an assay method with a PAR containing monoclonal anti-SV40.

SQ Sequence 44 BP; 15 A; 10 C; 5 G; 14 T;

Query Match 71.0%; Score 14.2; DB 1; Length 44;  
Best Local Similarity 84.2%; Pred. No. 90;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 ggaacttcctaaaggagg 19
    ||||| ||||| |||||
Db 28 GGAACATCCCTATAGTAG 10

RESULT 2
Q43300/c
ID Q43300 standard; DNA; 45 BP.
AC Q43300;
DE 23-SEP-1993 (first entry)
DE Sequence of PCR primer oligo 2 for the amplification of the
DE poliovirus genome from positions 1 to 740.
DE Recombinant virus; proteolytic cleavage site; vaccine;
KW exogenous nucleic acid; replication competent; ss.
OS Synthetic.
PN W09311251-A.
PD 10-JUN-1993.
PF 04-DEC-1992; U10543.
PR 06-DEC-1991; US-804893.
PR 18-SEP-1992; US-947790.
PA (AMCY ) AMERICAN CYANAMID CO.
PA (WHED ) WHITEHEAD INST BIOMEDICAL.
PI Andino R, Feinberg M, Reilly PA, Weeks-levy CB;
DR Recombinant viruses comprising artificial proteolytic cleavage
PT site - useful as vaccines against bacterial, viral and fungal
PT infections, parasitic diseases, cancer and allergies
PS Example; Table 1, page 36; 103pp; English.
CC There are a number of locations within the poliovirus genome at
CC which the exogenous nucleic acid sequence encoding the artificial
CC polypeptide and the nucleic acid sequences encoding the artificial
CC proteolytic cleavage sites can be positioned to produce replication-
CC competent recombinant polioviruses that express the encoded product.
CC These sites within the genome of the poliovirus include a terminal
CC end, the junction between the VP1 coding region and the 2A coding
CC region, the junction between the 2A coding region and the 2B coding
CC region and the junction between the 2C coding region and the 3A
CC coding region (see Q43294-98). Two independent PCR reactions were
CC performed using oligonucleotides 1 and 2, and 3 and 4 (see Q43299-
CC Q43302) to amplify portions of the poliovirus genome from positions
CC 1 to 740 and 740 to 1540, respectively. A second PCR amplification
CC was performed with oligonucleotide primers 1 and 4. The 1582 base
CC pair (bp) DNA fragment resulting from this amplification represents
CC the 42 nucleotide insert (carrying the novel restriction enzyme
CC sites and sequences encoding proteolytic processing signals) and
CC sequences 1 to 740, and 740 to 1540.
SQ Sequence 45 BP; 13 A; 13 C; 9 G; 10 T;

Query Match 68.0%; Score 13.6; DB 1; Length 45;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggagg 20
    ||||| ||||| |||||
Db 31 GGAATTCCTCGAGGAGG 12

RESULT 3
T73799
ID T73799 standard; DNA; 23 BP.
AC T73799;
DE 04-SEP-1997 (first entry)
DE Multiplex short-PCR primer for retinoblastoma gene exon 23.
DE Two dimensional gene scanning; TDGS; polymerase chain reaction; PCR;
KW electrophoresis; mutation; detection; simultaneous;
KW multiplex amplification; short distance PCR; ss.
OS Synthetic.
PN W09639535-A1.
PD 12-DEC-1996.
PF 03-JUN-1996; IB0543.
PR 06-JUN-1995; US-471249.
PA (LIDD/) LI D.

```

```

PA (VIJG/) VIJG J.
PI LI D, Vijg J;
DR WPI; 97-043156/04.
PT Detecting gene mutation(s) by two-step multiplex PCR amplification
PT - pref. followed by two-dimensional electrophoretic sepn. of
PT fragments on basis of size and base pair sequence
PS Disclosure; Page 17/1; 37pp; English.
CC Predetermined gene exons derived from DNA can be analysed by a new
CC method comprising: (a) adding primer pairs to successive groups of
CC the gene exons followed by PCR amplifications in a common tube, i.e.
CC multiplex long PCR; (b) adding further primer pairs specific to each
CC of the gene exons followed by PCR amplifications in a common tube,
CC i.e. multiplex short PCR; and (c) optionally electrophoretically
CC separating the gene fragments. Using the method, all possible mutations
CC in the gene(s) can be detected simultaneously based on a minimal number
CC of two-step multiplex PCR reactions in combination with automatic
CC two-dimensional separation of the fragments. In a specific example
CC of the new method, mutations in the retinoblastoma gene were
CC detected using 6 pairs of primers for the initial, long-PCR
CC followed by short-PCR amplifications using one pair of primers per
CC exon for exons 2-14 and 17-27. For the short-PCR, one primer in each
CC pair had a 5'-GC clamp. It was found that 5 ng of genomic DNA was
CC sufficient to obtain all the products and to allow detection of
CC mutations based on length polymorphisms.
SQ Sequence 23 BP; 9 A; 8 C; 2 G; 4 T;

Query Match 67.0%; Score 13.4; DB 1; Length 23;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 acttcctaaaggagg 18
    ||||| ||||| ||
Db 5 ACTTCCTAAGAGAGA 19

RESULT 4
X39216
ID X39216 standard; DNA; 31 BP.
AC X39216;
DE 15-JUN-1999 (first entry)
DE Human genomic DNA polymorphic site sequence tag 663.
DE Polymorphic site; human; forensic; paternity testing; phenotypic trait;
KW diagnosis; disease susceptibility; autoimmune disease; infection; cancer;
KW inflammatory disorder; nervous system disorder; longevity; drug response;
KW physical characteristic; therapy; breeding program; linkage; locus;
KW gene mapping; treatment; prevention; ss.
OS Homo sapiens.
PN W0914228-A1.
PD 25-MAR-1999.
PF 16-SEP-1998; U19325.
PR 18-NOV-1997; US-066172.
PR 17-SEP-1997; US-059304.
PA (AFFY-) AFFYMETRIX INC.
PI Berno A, Chee M, Fan J, Lipshutz RJ;
DR WPI; 99-229497/19.
PT Nucleic acid encoding specific human polymorphisms
PS Claim 1; Page 20; 56pp; English.
CC This invention describes nucleic acid segments represented in
CC X38554-X39408 which are isolated from any of about 750 human genomic
CC regions given in the specification that include a polymorphic site, or
CC their complements. Analysis of the polymorphisms is useful (1) to
CC identify individuals for forensic studies and paternity testing, (2) to
CC correlate the polymorphisms with phenotypic traits, e.g. for diagnosis
CC of, or susceptibility to, a wide range of diseases including autoimmune,
CC inflammatory and nervous system disorders, cancer, infections etc., also
CC longevity, physical characteristics, response to drugs or therapy, also
CC in animals and plants to identify individuals for breeding programs, (3)
CC to identify physical linkage between nucleic acid segments and a specific
CC genetic locus, associated with a trait for gene mapping and for
CC subsequent cloning of the gene responsible for the trait. The products of
CC the invention may also be used for treatment or prevention of the
CC specified diseases.

```

PF	20-DEC-1993; U12388.
PR	23-DEC-1992; US-996783.
PR	17-SEP-1993; US-123936.
PA	(GENE-) GENELABS TECHNOLOGIES INC.
PI	Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR	WFI; 94-234711/28.
PT	Sequence-directed DNA-binding molecules - useful in
PT	pharmaceuticals and as molecular reagents
PS	Claim 28; Page 277; 587pp; English.
CC	A DNA protein-binding assay is provided, useful for screening
CC	libraries of synthetic or biological cpds. for their ability
CC	to bind DNA test sequences. The assay is versatile in that any
CC	number of test sequences can be tested by placing the test sequence
CC	adjacent to a defined protein-binding screening sequence. Binding
CC	of mols. to these test sequences changes the binding characteristics
CC	of the protein mol. to its cognate binding sequence. When such a mol.
CC	binds the test sequence, the equilibrium of the DNA:protein complexes
CC	is disturbed, generating changes in the concentration of free DNA probe.
CC	One application of this method is to eucaryotic general transcription
CC	factors (e.g. TFIIID), where the target region is typically selected
CC	from DNA sequences adjacent to the binding site for the eucaryotic
CC	transcription factor. Numerous exemplary test sequences are given:
CC	the sequences in Q69251-731 and Q69850 correspond to promoter targets
CC	(typically, TATA box-contg. sites) for human genes and the sequences in
CC	Q69732-849 correspond to promoter targets for viral genes. The test
CC	sequences may also be randomly generated. DNA:protein interaction may
CC	be used for screening purposes, e.g. the Herpes Simplex Virus (HSV)
CC	origin of replication and UL9 (see Q69851-52, Q69865 and Q69891).
CC	Sequence 45 BP; 14 A; 9 C; 13 G; 9 T;
SQ	
Query Match	66.0%; Score 13.2; DB 1; Length 45;
Best Local Similarity	83.3%; Pred. No. 3e+02;
Matches	15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy	1 ggaacttccttaaggga 18 
Dd	12 GGAGCTTACATAAAGGGA 29
RESULT	7
Q69592	
ID	Q69592 standard; DNA; 45 BP.
AC	Q69592;
DE	01-MAR-1995 (first entry)
DT	Human gene for fibrinogen gamma chain target region.
KW	DNA protein-binding assay; test sequence; screening sequence;
KW	promoter; target; TATA box; Herpes Simplex Virus; HSV;
KW	origin of replication; UL9; transcription factor; TFIIID; ds.
OS	Synthetic.
PN	WQ9414980-A.
PD	07-JUL-1994.
PF	20-DEC-1993; U12388.
PR	23-DEC-1992; US-996783.
PR	17-SEP-1993; US-123936.
PA	(GENE-) GENELABS TECHNOLOGIES INC.
PI	Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR	WFI; 94-234711/28.
PT	Sequence-directed DNA-binding molecules - useful in
PT	pharmaceuticals and as molecular reagents
PS	Claim 28; Page 383; 587pp; English.
CC	A DNA protein-binding assay is provided, useful for screening
CC	libraries of synthetic or biological cpds. for their ability
CC	to bind DNA test sequences. The assay is versatile in that any
CC	number of test sequences can be tested by placing the test sequence
CC	adjacent to a defined protein-binding screening sequence. Binding
CC	of mols. to these test sequences changes the binding characteristics
CC	of the protein mol. to its cognate binding sequence. When such a mol.
CC	binds the test sequence, the equilibrium of the DNA:protein complexes
CC	is disturbed, generating changes in the concentration of free DNA probe.
CC	One application of this method is to eucaryotic general transcription
CC	factors (e.g. TFIIID), where the target region is typically selected
CC	from DNA sequences adjacent to the binding site for the eucaryotic

```
CC transcription factor. Numerous exemplary test sequences are given:
CC the sequences in Q69251-731 and Q69850 correspond to promoter targets
CC (typically, TATA box-contg. sites) for human genes and the sequences in
CC Q69732-849 correspond to promoter targets for viral genes. The test
CC sequences may also be randomly generated. DNA:protein interaction may
CC be used for screening purposes, e.g. the Herpes Simplex Virus (HSV)
CC origin of replication and UL9 (see Q69851-52, Q69865 and Q69891).
SQ Sequence 45 BP; 14 A; 9 C; 13 G; 9 T;

Query Match 66.0%; Score 13.2; DB 1; Length 45;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
    ||| ||| | |||||
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 8
T64054
ID T64054 standard; DNA; 45 BP.
AC T64054;
DT 14-MAR-1997 (first entry)
DE Human fibrinogen gamma chain gene TFIIID binding site.
KW Duplex DNA; target region; binding characteristic; DNA binding protein;
KW TFIIID; transcription factor; binding site; inhibition; enhance;
KW cancer; inherited genetic disorder; ds.
OS Homo sapiens.
PN US5578444-A.
PD 26-NOV-1996.
PF 27-JUN-1991; US-723618.
PR 23-DEC-1992; US-996783.
PR 17-SEP-1993; US-123936.
PR 20-DEC-1993; US-171389.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR WPI; 97-020402/02.
PT Altering binding characteristics of DNA binding proteins to duplex
PT DNA - by attaching specific small cpd. to target region close to the
PT protein's binding site, useful in treatment of viral disease, cancer
PT etc
PS Claim 6; Column 273; 264pp; English.
CC The sequences given in T63713-4312 represent duplex DNA's which act
CC as target regions in the method of the invention. The method for
CC altering the binding characteristics of a DNA-binding protein to duplex
CC DNA comprises contacting the duplex DNA with a small molecule which
CC binds sequence-specifically to a target region, where, when the small
CC molecule is bound to the target region, it is adjacent to, but not
CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.
CC The small molecule is added at a concentration effective to alter the
CC binding of the DNA binding protein, pref. TFIIID, to its binding site on
CC the duplex DNA. The binding of the small molecule may inhibit or enhance
CC compounds isolated using this method are potentially useful as
CC therapeutic agents for treatment of any disease which involves a
CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.
CC The method is suitable for screening large biological or chemical
CC libraries and allows determination of sequence-specific and relative
CC affinities of known DNA-binding agents for different DNA sequences.
CC The design of these duplex DNA's allows a single DNA:protein interaction
CC to be used for screening sequence-specific, or preferential, DNA binding
CC proteins that recognise almost any possible sequence (see also T49539-
CC 74).
SQ Sequence 45 BP; 14 A; 9 C; 13 G; 9 T;

Query Match 66.0%; Score 13.2; DB 1; Length 45;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
    ||| ||| | |||||
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 8
T64054
ID T64054 standard; DNA; 45 BP.
AC T64054;
DT 14-MAR-1997 (first entry)
DE Human fibrinogen gamma chain gene TFIIID binding site.
KW Duplex DNA; target region; binding characteristic; DNA binding protein;
KW TFIIID; transcription factor; binding site; inhibition; enhance;
KW cancer; inherited genetic disorder; ds.
OS Homo sapiens.
PN US5578444-A.
PD 26-NOV-1996.
PF 27-JUN-1991; US-723618.
PR 23-DEC-1992; US-996783.
PR 17-SEP-1993; US-123936.
PR 20-DEC-1993; US-171389.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR WPI; 97-020402/02.
PT Altering binding characteristics of DNA binding proteins to duplex
PT DNA - by attaching specific small cpd. to target region close to the
PT protein's binding site, useful in treatment of viral disease, cancer
PT etc
PS Claim 6; Column 273; 264pp; English.
CC The sequences given in T63713-4312 represent duplex DNA's which act
CC as target regions in the method of the invention. The method for
CC altering the binding characteristics of a DNA-binding protein to duplex
CC DNA comprises contacting the duplex DNA with a small molecule which
CC binds sequence-specifically to a target region, where, when the small
CC molecule is bound to the target region, it is adjacent to, but not
CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.
CC The small molecule is added at a concentration effective to alter the
CC binding of the DNA binding protein, pref. TFIIID, to its binding site on
CC the duplex DNA. The binding of the small molecule may inhibit or enhance
CC compounds isolated using this method are potentially useful as
CC therapeutic agents for treatment of any disease which involves a
CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.
CC The method is suitable for screening large biological or chemical
CC libraries and allows determination of sequence-specific and relative
CC affinities of known DNA-binding agents for different DNA sequences.
CC The design of these duplex DNA's allows a single DNA:protein interaction
CC to be used for screening sequence-specific, or preferential, DNA binding
CC proteins that recognise almost any possible sequence (see also T49539-
CC 74).
SQ Sequence 45 BP; 14 A; 9 C; 13 G; 9 T;

Query Match 66.0%; Score 13.2; DB 1; Length 45;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
    ||| ||| | |||||
Db 12 GGAGCTTACATAAAGGGA 29
```

```
||| ||| | |||||
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 9
T63842
ID T63842 standard; DNA; 45 BP.
AC T63842;
DT 13-MAR-1997 (first entry)
DE Human fibrinogen gamma chain gene TFIIID binding site.
KW Duplex DNA; target region; binding characteristic; DNA binding protein;
KW TFIIID; transcription factor; binding site; inhibition; enhance;
KW cancer; inherited genetic disorder; ds.
OS Homo sapiens.
PN US5578444-A.
PD 26-NOV-1996.
PF 27-JUN-1991; US-723618.
PR 23-DEC-1992; US-996783.
PR 17-SEP-1993; US-123936.
PR 20-DEC-1993; US-171389.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR WPI; 97-020402/02.
PT Altering binding characteristics of DNA binding proteins to duplex
PT DNA - by attaching specific small cpd. to target region close to the
PT protein's binding site, useful in treatment of viral disease, cancer
PT etc
PS Claim 6; Column 165-166; 264pp; English.
CC The sequences given in T63713-4312 represent duplex DNA's which act
CC as target regions in the method of the invention. The method for
CC altering the binding characteristics of a DNA-binding protein to duplex
CC DNA comprises contacting the duplex DNA with a small molecule which
CC binds sequence-specifically to a target region, where, when the small
CC molecule is bound to the target region, it is adjacent to, but not
CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.
CC The small molecule is added at a concentration effective to alter the
CC binding of the DNA binding protein, pref. TFIIID, to its binding site on
CC the duplex DNA. The binding of the small molecule may inhibit or enhance
CC compounds isolated using this method are potentially useful as
CC therapeutic agents for treatment of any disease which involves a
CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.
CC The method is suitable for screening large biological or chemical
CC libraries and allows determination of sequence-specific and relative
CC affinities of known DNA-binding agents for different DNA sequences.
CC The design of these duplex DNA's allows a single DNA:protein interaction
CC to be used for screening sequence-specific, or preferential, DNA binding
CC proteins that recognise almost any possible sequence (see also T49539-
CC 74).
SQ Sequence 45 BP; 14 A; 9 C; 13 G; 9 T;

Query Match 66.0%; Score 13.2; DB 1; Length 45;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggga 18
    ||| ||| | |||||
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 10
X17342
ID X17342 standard; DNA; 45 BP.
AC X17342;
DT 06-MAY-1999 (first entry)
DE Test sequence from human gene for fibrinogen gamma chain.
KW Test sequence; DNA-binding molecule; screening sequence; human;
KW nucleic acid amplification; target; viral; ds.
OS Homo sapiens.
PN US5869241-A.
PD 09-FEB-1999.
```

PF 07-JUN-1995; 475228.  
 PR 20-DEC-1993; US-171389.  
 PR 27-JUN-1991; US-723618.  
 PR 23-DEC-1992; US-996783.  
 PR 17-SEP-1993; US-123936.  
 PR 07-JUN-1995; US-475228.  
 PA (GENE-) GENELABS TECHNOLOGIES INC.  
 PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;  
 WIPI: 99-152755/13.  
 PT Determination of DNA sequence preference of a DNA-binding molecule -  
 PT based on inhibition of binding of protein to oligonucleotide  
 PT sequence attached to test sequence  
 PS Claim 3; Columns 275-276; 270pp; English.  
 CC Sequences X17001 to X17600 represent specifically claimed target test  
 CC sequences that are used in the method of the invention of determining the  
 CC DNA sequence preference of a DNA-binding molecule. The method comprises:  
 CC (i) adding a test molecule and a DNA-binding protein to a mixture of  
 CC duplex DNA test oligonucleotides, each of the test oligonucleotides  
 CC having a test sequence adjacent to a screening sequence, where the  
 CC screening sequence binds to the DNA-binding protein with a binding  
 CC affinity that is independent of the DNA sequence of the test sequence,  
 CC and where the mixture of duplex DNA test oligonucleotides includes  
 CC several test sequences; (ii) incubating the test molecule, the mixture of  
 CC duplex DNA test oligonucleotides and the DNA-binding protein for a time  
 CC sufficient to permit binding of the test molecule to test sequences in  
 CC the duplex DNA; (iii) separating unbound test oligonucleotides from test  
 CC oligonucleotides bound to binding protein; (iv) amplifying the unbound  
 CC test oligonucleotides; (v) repeating steps (ii) to (iv); (vi) isolating  
 CC the amplified test oligonucleotides; and (vii) sequencing the isolated  
 CC test oligonucleotides. Test sequences X17001-X17481 and X17600 correspond  
 CC to promoter targets for human genes and test sequences X17482-X17599  
 CC correspond to promoter targets for viral genes.  
 SQ Sequence 45 BP; 14 A; 9 C; 13 G; 9 T;

Query Match 66.0%; Score 13.2; DB 1; Length 45;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaacttcctctaaaggga 18  
 ||| ||| | |||||  
 DB 12 GGAGCTTACATAAAGGGA 29

RESULT 11  
 X17130  
 ID X17130 standard; DNA; 45 BP.  
 AC X17130;  
 DT 06-MAY-1999 (first entry)  
 DE Test sequence from human fibrinogen gamma chain gene.  
 KW Test sequence: DNA-binding molecule; screening sequence; human;  
 KW nucleic acid amplification; target; viral; ds.  
 OS Homo sapiens.  
 PN US5869241-A.  
 PD 09-FEB-1999.  
 PF 07-JUN-1995; 475228.  
 PR 20-DEC-1993; US-171389.  
 PR 27-JUN-1991; US-723618.  
 PR 23-DEC-1992; US-996783.  
 PR 17-SEP-1993; US-123936.  
 PR 07-JUN-1995; US-475228.  
 PA (GENE-) GENELABS TECHNOLOGIES INC.  
 PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;  
 WIPI: 99-152755/13.  
 PT Determination of DNA sequence preference of a DNA-binding molecule -  
 PT based on inhibition of binding of protein to oligonucleotide  
 PT sequence attached to test sequence  
 PS Claim 3; Columns 167-168; 270pp; English.  
 CC Sequences X17001 to X17600 represent specifically claimed target test  
 CC sequences that are used in the method of the invention of determining the  
 CC DNA sequence preference of a DNA-binding molecule. The method comprises:  
 CC (i) adding a test molecule and a DNA-binding protein to a mixture of  
 CC duplex DNA test oligonucleotides, each of the test oligonucleotides

CC having a test sequence adjacent to a screening sequence, where the  
 CC screening sequence binds to the DNA-binding protein with a binding  
 CC affinity that is independent of the DNA sequence of the test sequence,  
 CC and where the mixture of duplex DNA test oligonucleotides includes  
 CC several test sequences; (ii) incubating the test molecule, the mixture of  
 CC duplex DNA test oligonucleotides and the DNA-binding protein for a time  
 CC sufficient to permit binding of the test molecule to test sequences in  
 CC the duplex DNA; (iii) separating unbound test oligonucleotides from test  
 CC oligonucleotides bound to binding protein; (iv) amplifying the unbound  
 CC test oligonucleotides; (v) repeating steps (ii) to (iv); (vi) isolating  
 CC the amplified test oligonucleotides; and (vii) sequencing the isolated  
 CC test oligonucleotides. Test sequences X17001-X17481 and X17600 correspond  
 CC to promoter targets for human genes and test sequences X17482-X17599  
 CC correspond to promoter targets for viral genes.  
 SQ Sequence 45 BP; 14 A; 9 C; 13 G; 9 T;

Query Match 66.0%; Score 13.2; DB 1; Length 45;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaacttcctctaaaggga 18  
 ||| ||| | |||||  
 DB 12 GGAGCTTACATAAAGGGA 29

RESULT 12  
 X36364  
 ID X36364 standard; DNA; 46 BP.  
 AC X36364;  
 DT 01-JUL-1999 (first entry)  
 DE Primer for human TIE ligand NL8 coding sequence.  
 KW Human TIE ligand; NL4; NL5; NL8; Ig homology domain; angiogenesis;  
 KW EGF homology domain; receptor tyrosine kinase; vascular endothelial cell;  
 KW early haemopoietic cell; receptor tyrosine kinase; neovascularisation; wound healing;  
 KW endothelial cell; growth inhibitor; apoptosis inducer; tumour cell;  
 KW vasculogenesis; detection; diagnosis; therapy; primer; ss.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO915653-A2.  
 PD 01-APR-1999.  
 PF 14-SEP-1998; U19093.  
 PR 29-OCT-1997; US-960507.  
 PR 19-SEP-1997; US-933821.  
 PA (GETH ) GENENTECH INC.  
 PI Botstein D, Ferrara N, Goddard A, Godowski PJ, Gurney AL,  
 PI Hillan K, Roy M, Schwall R, Tumas D;  
 WIPI: 99-263480/22.  
 DT New isolated TIE ligand homologs for, e.g. developing products for  
 PT treatment of tumors  
 PT Example 4; Page 54; 132pp; English.  
 CC This sequence is a primer for DNA encoding the human tyrosine kinase  
 CC containing Ig and EGF homology domains (TIE) ligand of the invention,  
 CC designated NL8. The TIE receptors are receptor tyrosine kinases which  
 CC are expressed in vascular endothelial cells and early haemopoietic cells.  
 CC The TIE receptors are believed to be actively involved in angiogenesis,  
 CC and may play a role in haemopoiesis as well. The TIE ligand homologs can  
 CC promote the survival and/or growth and/or differentiation of TIE receptor  
 CC expressing cells. They can be used for promoting neovascularisation in  
 CC wound healing and for promoting angiogenic processes, such as for  
 CC inducing collateral vascularisation in an ischaemic heart or limb, or for  
 CC promoting bone development and/or maturation and/or growth in a patient  
 CC or muscle growth and development. The TIE ligand homologs and antibodies  
 CC can inhibit the growth of endothelial cells and induce apoptosis of  
 CC cells, particularly tumour cells. They can inhibit vasculogenesis,  
 CC particularly the vascularisation of tumour cells. The antibodies can also  
 CC inhibit vascularisation of a cell in which a gene encoding an NL1, NL5,  
 CC NL8 or NL4 polypeptide is amplified. The products can also be used for  
 CC detection, diagnosis, drug screening and production of transgenic  
 CC animals.  
 SQ Sequence 46 BP; 11 A; 10 C; 15 G; 10 T;

```

Query Match      66.0%; Score 13.2; DB 1; Length 46;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aacttcctaaaggagg 20
   | | | | | | | | | |
Db 12 ACCCTCACTAAGGAGG 29

RESULT 13
X36360
ID X36360 standard; DNA; 48 BP.
AC X36360;
DT 01-JUL-1999 (first entry)
DE Primer for human TIE ligand NL5 coding sequence.
KW Human TIE ligand; NL1; NL4; NL5; NL8; Ig homology domain; angiogenesis;
KW EGF homology domain; receptor tyrosine kinase; vascular endothelial cell;
KW early haemopoietic cell; haemopoiesis; neovascularisation; wound healing;
KW endothelial cell; growth inhibitor; apoptosis inducer; tumour cell;
KW vasculogenesis; detection; diagnosis; therapy; primer; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9915653-A2.
PD 01-APR-1999.
PF 14-SEP-1998; U19093.
PR 29-OCT-1997; US-960507.
PR 19-SEP-1997; US-933821.
PA (GETH ) GENENTECH INC.
PI Botstein D, Ferrara N, Goddard A, Godowski PJ, Gurney AL,
PI Hillan K, Roy M, Schwall R, Tumas D;
PT WPI; 99-263480/22.
PT New isolated TIE ligand homologs for, e.g. developing products for
PT treatment of tumors
PS Example 4; Page 53; 132pp; English.
CC This sequence is a primer for DNA encoding the human tyrosine kinase
CC containing Ig and EGF homology domains (TIE) ligand of the invention,
CC designated NL5. The TIE receptors are receptor tyrosine kinases which
CC are expressed in vascular endothelial cells and early haemopoietic cells.
CC The TIE receptors are believed to be actively involved in angiogenesis,
CC and may play a role in haemopoiesis as well. The TIE ligand homologs can
CC promote the survival and/or growth and/or differentiation of TIE receptor
CC expressing cells. They can be used for promoting neovascularisation in
CC wound healing and for promoting angiogenic processes, such as for
CC inducing collateral vascularisation in an ischaemic heart or limb, or for
CC promoting bone development and/or maturation and/or growth in a patient
CC or muscle growth and development. The TIE ligand homologs and antibodies
CC can inhibit the growth of endothelial cells and induce apoptosis of
CC cells, particularly tumour cells. They can inhibit vasculogenesis,
CC particularly the vascularisation of tumour cells. The antibodies can also
CC inhibit vascularisation of a cell in which a gene encoding an NL1, NL5,
CC NL8 or NL4 polypeptide is amplified. The products can also be used for
CC detection, diagnosis, drug screening and production of transgenic
CC animals.
SQ Sequence 48 BP; 14 A; 7 C; 14 G; 13 T;

Query Match      66.0%; Score 13.2; DB 1; Length 48;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aacttcctaaaggagg 20
   | | | | | | | | | |
Db 12 ACCCTCACTAAGGAGG 29

RESULT 14
X52520
ID X52520 standard; DNA; 48 BP.
AC X52520;
DT 25-JUN-1999 (first entry)
DE PCR primer for in situ analysis of DNA5918-1174.
KW Secreted protein; transmembrane protein; human; enterocolitis;
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;

```

```

KW congenital microvillus atrophy; skin disease; cell growth;
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW anti-thrombotic; wound healing; tissue repair; PCR primer; ss.
OS Synthetic.
PN WO9914328-A2.
PD 25-MAR-1999.
PF 16-SEP-1998; U19330.
PR 25-NOV-1997; US-066840.
PR 17-SEP-1997; US-059113.
PR 17-SEP-1997; US-059115.
PR 17-SEP-1997; US-059117.
PR 17-SEP-1997; US-059119.
PR 17-SEP-1997; US-059121.
PR 17-SEP-1997; US-059122.
PR 17-SEP-1997; US-059184.
PR 18-SEP-1997; US-059263.
PR 18-SEP-1997; US-059266.
PR 15-OCT-1997; US-062125.
PR 17-OCT-1997; US-062285.
PR 17-OCT-1997; US-062287.
PR 21-OCT-1997; US-063486.
PR 24-OCT-1997; US-062814.
PR 24-OCT-1997; US-062816.
PR 24-OCT-1997; US-063045.
PR 24-OCT-1997; US-063120.
PR 24-OCT-1997; US-063121.
PR 24-OCT-1997; US-063127.
PR 24-OCT-1997; US-063128.
PR 27-OCT-1997; US-063329.
PR 27-OCT-1997; US-063327.
PR 28-OCT-1997; US-063541.
PR 28-OCT-1997; US-063542.
PR 28-OCT-1997; US-063544.
PR 28-OCT-1997; US-063549.
PR 28-OCT-1997; US-063550.
PR 28-OCT-1997; US-063564.
PR 29-OCT-1997; US-063435.
PR 29-OCT-1997; US-063704.
PR 29-OCT-1997; US-063732.
PR 29-OCT-1997; US-063738.
PR 29-OCT-1997; US-063734.
PR 29-OCT-1997; US-064215.
PR 29-OCT-1997; US-063735.
PR 31-OCT-1997; US-063870.
PR 31-OCT-1997; US-064103.
PR 03-NOV-1997; US-064248.
PR 07-NOV-1997; US-064809.
PR 12-NOV-1997; US-065186.
PR 17-NOV-1997; US-065846.
PR 18-NOV-1997; US-065693.
PR 21-NOV-1997; US-066120.
PR 21-NOV-1997; US-066364.
PR 24-NOV-1997; US-066772.
PR 24-NOV-1997; US-066466.
PR 24-NOV-1997; US-066770.
PR 24-NOV-1997; US-066511.
PR 24-NOV-1997; US-066453.
PA (GETH ) GENENTECH INC.
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
DR WPI; 99-229533/19.
PT New isolated human genes and polypeptides used in, e.g. treatment of
PT gastrointestinal ulceration
PS Example 74; Page 179; 320pp; English.
CC Oligonucleotides X52276-532 represent PCR primers and probes used
CC to isolate and amplify cDNA encoding secreted and transmembrane human
CC proteins (see X52213-74 and Y13344-403). The cDNA sequences are
CC obtained from cDNA libraries, prepared from fetal lung, fetal kidney,
CC fetal brain, fetal liver and fetal retina. The encoded polypeptides
CC have specific uses based on their homology to known polypeptides,
CC e.g. PRO211 and PRO217 can be used for disorders associated with the
CC preservation and maintenance of gastrointestinal mucosa and the

```

CC repair of acute and chronic mucosal lesions (e.g. enterocolitis,  
 CC Zollinger-Ellison syndrome, gastrointestinal ulceration and congenital  
 CC microvillus atrophy), skin diseases associated with abnormal keratinocyte  
 CC differentiation (e.g. psoriasis, epithelial cancers such as lung squamous  
 CC cell carcinoma of the vulva and gliomas), potent effects on cell growth  
 CC and development, diseases related to growth or survival of nerve cells  
 CC including Parkinson's disease, Alzheimer's disease, ALS, neuropathies or  
 CC cancer. PRO265 can be used as for fibromodulin, e.g. for reducing dermal  
 CC scarring. PRO264 can be used as a target for anti-tumor drugs. PRO533  
 CC may be used in the treatment of Usher Syndrome or Atrophia areata;  
 CC PRO269 can be used as an anti-thrombotic agent; PRO287 polypeptides  
 CC and portions may have therapeutic applications in wound healing and  
 CC tissue repair; PRO317 can be used for treating problems of the kidney,  
 CC uterus, endometrium, blood vessels, or related tissue, e.g. in the  
 CC heart of genital tract.  
 SQ Sequence 48 BP; 15 A; 14 C; 10 G; 9 T;

Query Match 66.0%; Score 13.2; DB 1; Length 48;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aacttcctaaaggagg 20  
 | | | | | | | | | |  
 DB 12 ACCCTCACTAAAGGGAGG 29

## RESULT 15

X52518  
 ID X52518 standard; DNA; 48 BP.

AC X52518; (first entry)  
 DT PCR primer for in situ analysis of DNA33089-1132.  
 DE Secreted protein; transmembrane protein; human; enterocolitis;  
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
 KW congenital microvillus atrophy; skin disease; cell growth;  
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
 KW anti-thrombotic; wound healing; tissue repair; PCR primer; ss.  
 OS Synthetic.  
 PN WO9914328-A2.

PD 25-MAR-1999.  
 PF 16-SEP-1998; U19330.

PR 25-NOV-1997; US-066840.  
 PR 17-SEP-1997; US-059113.

PR 17-SEP-1997; US-059115.  
 PR 17-SEP-1997; US-059117.

PR 17-SEP-1997; US-059119.  
 PR 17-SEP-1997; US-059121.

PR 17-SEP-1997; US-059122.  
 PR 18-SEP-1997; US-059184.

PR 18-SEP-1997; US-059263.  
 PR 18-SEP-1997; US-059266.

PR 15-OCT-1997; US-062125.  
 PR 17-OCT-1997; US-062285.

PR 17-OCT-1997; US-062287.  
 PR 21-OCT-1997; US-063486.

PR 24-OCT-1997; US-062814.  
 PR 24-OCT-1997; US-062816.

PR 24-OCT-1997; US-063045.  
 PR 24-OCT-1997; US-063120.

PR 24-OCT-1997; US-063121.  
 PR 24-OCT-1997; US-063127.

PR 24-OCT-1997; US-063128.  
 PR 27-OCT-1997; US-063329.

PR 27-OCT-1997; US-063327.  
 PR 28-OCT-1997; US-063341.

PR 28-OCT-1997; US-063542.  
 PR 28-OCT-1997; US-063544.

PR 28-OCT-1997; US-063549.  
 PR 28-OCT-1997; US-063550.

PR 28-OCT-1997; US-063564.  
 PR 28-OCT-1997; US-063564.

PR 29-OCT-1997; US-063435.  
 PR 29-OCT-1997; US-063704.  
 PR 29-OCT-1997; US-063732.  
 PR 29-OCT-1997; US-063738.  
 PR 29-OCT-1997; US-063734.  
 PR 29-OCT-1997; US-064215.  
 PR 29-OCT-1997; US-063735.  
 PR 31-OCT-1997; US-063870.  
 PR 31-OCT-1997; US-064103.  
 PR 03-NOV-1997; US-064248.  
 PR 07-NOV-1997; US-064809.  
 PR 12-NOV-1997; US-065186.  
 PR 17-NOV-1997; US-065846.  
 PR 18-NOV-1997; US-065693.  
 PR 21-NOV-1997; US-066120.  
 PR 21-NOV-1997; US-066364.  
 PR 24-NOV-1997; US-066772.  
 PR 24-NOV-1997; US-066466.  
 PR 24-NOV-1997; US-066770.  
 PR 24-NOV-1997; US-066511.  
 PR 24-NOV-1997; US-066453.  
 PA (GETH) GENENTECH INC.  
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WL, Yuan J;  
 PT WPI; 99-229533/19.  
 DR New isolated human genes and polypeptides used in, e.g. treatment of  
 PT gastrointestinal ulceration  
 PS Example 74; Page 179; 320pp; English.

CC Oligonucleotides X52276-532 represent PCR primers and probes used  
 CC to isolate and amplify cDNA encoding secreted and transmembrane human  
 CC proteins (see X52213-74 and Y1344-403). The cDNA sequences are  
 CC obtained from cDNA libraries, prepared from fetal lung, fetal kidney,  
 CC fetal brain, fetal liver and fetal retina. The encoded polypeptides  
 CC have specific uses based on their homology to known polypeptides, the  
 CC e.g. PRO211 and PRO217 can be used for disorders associated with the  
 CC preservation and maintenance of gastrointestinal mucosa and the  
 CC repair of acute and chronic mucosal lesions (e.g. enterocolitis,  
 CC Zollinger-Ellison syndrome, gastrointestinal ulceration and congenital  
 CC microvillus atrophy), skin diseases associated with abnormal keratinocyte  
 CC differentiation (e.g. psoriasis, epithelial cancers such as lung squamous  
 CC cell carcinoma of the vulva and gliomas), potent effects on cell growth  
 CC and development, diseases related to growth or survival of nerve cells  
 CC including Parkinson's disease, Alzheimer's disease, ALS, neuropathies or  
 CC cancer. PRO265 can be used as for fibromodulin, e.g. for reducing dermal  
 CC scarring. PRO264 can be used as a target for anti-tumor drugs. PRO533  
 CC may be used in the treatment of Usher Syndrome or Atrophia areata;  
 CC PRO269 can be used as an anti-thrombotic agent; PRO287 polypeptides  
 CC and portions may have therapeutic applications in wound healing and  
 CC tissue repair; PRO317 can be used for treating problems of the kidney,  
 CC uterus, endometrium, blood vessels, or related tissue, e.g. in the  
 CC heart of genital tract.  
 SQ Sequence 48 BP; 17 A; 7 C; 13 G; 11 T;

Query Match 66.0%; Score 13.2; DB 1; Length 48;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aacttcctaaaggagg 20  
 | | | | | | | | | |  
 DB 12 ACCCTCACTAAAGGGAGG 29

Search completed: May 23, 2000, 11:21:35  
 Job time: 6017 sec

---

GenCore version 4.5 Copyright (c) 1993 - 2000 CompuGen Ltd.	
OM nucleic - nucleic search, using sw model	
Run on:	May 23, 2000, 11:00:22 ; Search time 1337.41 Seconds (without alignments) 60.613 Million cell updates/sec
Title:	US-08-945-805-3
Perfect score:	20
Sequence:	1 ggaacttcctaaaggagg 20
Scoring table:	IDENTITY_NUC Gapop 10.0 , Gapext 1.0
Searched:	4857316 seqs, 2026611650 residues
Total number of hits satisfying chosen parameters:	23194
Minimum DB seq length: 0 Maximum DB seq length: 50	
Post-processing: Minimum Match 0% Listing first 45 summaries	
Database :	EST:* 1: em_est1:* 2: em_est2:* 3: em_est3:* 4: em_est4:* 5: em_est5:* 6: em_est6:* 7: em_est7:* 8: em_est8:* 9: em_est9:* 10: em_est10:* 11: em_est11:* 12: em_est12:* 13: em_est13:* 14: em_est14:* 15: em_est15:* 16: em_est16:* 17: em_est17:* 18: em_est18:* 19: em_est19:* 20: gb_est1:* 21: gb_est2:* 22: gb_est3:* 23: gb_est4:* 24: gb_est5:* 25: gb_est6:* 26: gb_est7:* 27: gb_est8:* 28: gb_est9:* 29: gb_est10:* 30: gb_est11:* 31: gb_est12:* 32: gb_est13:* 33: gb_est14:* 34: gb_est15:* 35: gb_est16:* 36: gb_est17:* 37: gb_est18:* 38: gb_est19:* 39: gb_est20:* 40: gb_est21:* 41: gb_est22:* 42: gb_est23:* 43: gb_est24:* 44: gb_est25:*
	45: gb_est26:* 46: gb_est27:* 47: gb_est28:* 48: gb_est29:* 49: gb_est30:* 50: gb_est31:* 51: gb_est32:* 52: em_est20:* 53: em_est21:* 54: em_est22:* 55: em_est23:* 56: em_est24:* 57: em_est25:* 58: em_est26:* 59: gb_est33:* 60: gb_est34:* 61: gb_est35:* 62: gb_est36:* 63: gb_est37:* 64: gb_est38:* 65: em_est27:* 66: em_est28:* 67: em_est29:* 68: em_est30:* 69: gb_est39:* 70: gb_est40:* 71: gb_est41:* 72: gb_est42:* 73: gb_est43:* 74: gb_est44:* 75: em_est31:* 76: em_est32:* 77: em_est33:* 78: em_est34:* 79: gb_est45:* 80: gb_est46:* 81: gb_est47:* 82: gb_gss1:* 83: gb_gss2:* 84: gb_gss3:* 85: gb_gss4:* 86: em_gss1:* 87: em_gss2:* 88: em_gss3:* 89: em_gss4:* 90: gb_gss5:* 91: gb_gss6:* 92: gb_gss7:* 93: gb_gss8:* 94: gb_gss9:* 95: em_gss5:* 96: em_gss6:* 97: em_gss7:* 98: em_gss8:* 99: em_gss9:* 100: em_gss10:* 101: em_gss11:* 102: gb_gss10:* 103: gb_gss11:* 104: em_gss12:* 105: gb_gss12:* 106: gb_gss13:* 107: gb_gss14:* 108: gb_gss15:* 109: gb_gss16:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.	
Result	% Query
SUMMARIES	

No.	Score	Match	Length	DB	ID	Description
c 1	11.8	59.0	46	25	N44913	N44913 yy21h06.r1
2	11.6	58.0	43	49	AI654728	AI654728 tq93e01.x
3	11.6	58.0	43	63	AI943364	AI943364 fc79c07.y
4	11.6	58.0	46	23	H55376	H55376 CHR220315 C
c 5	11.6	58.0	46	36	AA613699	AA613699 no38f08.s
c 6	11.6	58.0	46	47	AI544460	AI544460 fb75b10.x
7	11.4	57.0	46	49	AI631068	AI631068 tz32b08.x
8	11.2	56.0	31	49	AI649395	AI649395 uk27h02.x
9	11.2	56.0	34	43	AI186043	AI186043 qe50f03.x
c 10	11.2	56.0	43	23	H40524	H40524 yp63f01.s1
c 11	11.2	56.0	49	22	R70689	R70689 vi41e11.s1
12	11.2	56.0	50	81	AJ281329	AJ281329 4A3A-P2A7
c 13	11.1	55.0	40	60	AI788409	AI788409 uk58d09.y
14	11.1	55.0	43	36	AA614588	AA614588 np50d01.s
15	11.1	55.0	46	33	AA411090	AA411090 zt36g09.r
c 16	11.1	55.0	49	41	AI036013	AI036013 vz68a10.r
17	10.8	54.0	34	36	AA624893	AA624893 vn83h01.r
c 18	10.8	54.0	37	35	AA594384	AA594384 nl93e10.s
19	10.8	54.0	39	60	AI811713	AI811713 tw75b03.x
20	10.8	54.0	43	46	AI440006	AI440006 tl69g11.x
21	10.8	54.0	50	43	AI219842	AI219842 q983d08.x
c 22	10.6	53.0	31	48	AI587478	AI587478 tr51c05.x
23	10.6	53.0	36	79	AW246469	AW246469 2821791.3
c 24	10.6	53.0	40	28	AA112852	AA112852 zm63c10.s
c 25	10.6	53.0	42	23	HI9551	HI9551 yn54f08.r1
c 26	10.6	53.0	42	45	AJ237207	AJ237207 AJ237207
27	10.6	53.0	46	48	AI582875	AI582875 ts07a07.x
28	10.6	53.0	50	22	R53613	R53613 yj71e05.s1
29	10.4	52.0	34	40	AA996016	AA996016 os26d10.s
30	10.4	52.0	41	22	D38738	D38738 HUMCIEF81B
31	10.4	52.0	44	24	D67713	D67713 CELK076H2F
c 32	10.4	52.0	49	42	AI095777	AI095777 qp30d06.x
33	10.4	52.0	49	43	AI244893	AI244893 qj98f06.x
34	10.4	52.0	49	43	AI244893	AI244893 qj98f06.x
35	10.2	51.0	40	33	AA416473	AA416473 vd11f11.s
c 36	10.2	51.0	41	22	D38738	D38738 HUMCIEF81B
37	10.2	51.0	44	26	AA600013	AA600013 ag29h11.s
c 38	10.2	51.0	46	24	DI9556	DI9556 MUSGS00956
39	10.2	51.0	46	48	AI569880	AI569880 tr57a07.x
c 40	10.2	51.0	49	44	AI273069	AI273069 qv62g02.x
c 41	10.2	51.0	50	81	AJ281329	AJ281329 4A3A-P2A7
42	10.0	50.0	34	45	AI351102	AI351102 qt23b06.x
c 43	10.0	50.0	36	79	AW246491	AW246491 2821603.3
c 44	10.0	50.0	36	103	AQ254660	AQ254660 EP(3)0881
45	10.0	50.0	37	45	AI351639	AI351639 qr06g08.x

## ALIGNMENTS

RESULT	1
LOCUS	N44913/c
DEFINITION	yy21h06.r1 Soares melanocyte 2NBHM Homo sapiens cdna clone IMAGE:271931 5' similar to gb:X57025_rnal INSULIN-LIKE GROWTH FACTOR 1A PRECURSOR (HUMAN); mRNA sequence.
ACCESSION	N44913
VERSION	N44913.1 GI:1186079
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 46) Hillier, L., Clark, N., Dubuc, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)

COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: T7 High quality sequence stop: 10.
FEATURES	Location/Qualifiers 1..46 source
BASE COUNT	5 a 14 c 11 g 16 t
ORIGIN	/organism="Homo sapiens" /db_xref="GDB:3881573" /db_xref="taxon:9606" /clone="IMAGE:271931" /clone_lib="Soares melanocyte 2NBHM" /sex="Male" /tissue_type="melanocyte" /lab_host="DH10B (ampicillin resistant)" /note="Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGCGGCGCGAGTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."
Query Match	59.0%; Score 11.8; DB 25; Length 46;
Best Local Similarity	86.7%; Pred. No. 2.2e+04;
Matches	13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ggaacttcctaaag 15 
Db	22 GGAACCTTGCCCAAG 8
RESULT	2
LOCUS	AI654728
DEFINITION	tq93e01.x1 NCI-CGAP_Ov23 Homo sapiens cdna clone IMAGE:2216376 3' similar to SW:YM16_MARPO P38459 HYPOTHETICAL 29.4 KD PROTEIN IN NAD4L-COX2 INTERGENIC REGION ;contains element TAR1 repetitive element ; mRNA sequence.
ACCESSION	AI654728
VERSION	AI654728.1 GI:4738707
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 43) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On May 18, 1998 this sequence version replaced gi:3138496. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.

# FEATURES

source  
Location/Qualifiers  
1. .43  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2216376"  
/clone\_lib="NCI\_CGAP\_Ov23"  
/tissue\_type="tumor, 5 pooled (see description)"  
/lab\_host="DH10B"  
/note="Organ: ovary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.35 kb. Tumor types include: mixed  
Mullerian tumor, papillary serous, clear cell, spindle  
cell. All are primary tumors, metastasis positive. Life  
Technologies catalog #: 11534-013"  
25 a 7 c 11 g 0 t  
BASE COUNT  
ORIGIN  
Query Match 58.0%; Score 11.6; DB 49; Length 43;  
Best Local Similarity 77.8%; Pred. No. 2.8e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 3 aacttcctaaaggagg 20  
||| ||| ||| ||| ||| |||  
Db 25 AACACCCCAAGGAAG 42  
RESULT 3  
LOCUS AI943364 43 bp mRNA EST 06-AUG-1999  
DEFINITION fc79c07.y1 zebrafish WASHU MPING EST Danio rerio cDNA 5' similar to  
tr:Q91594 Q91594 CELLULAR NUCLEIC ACID BINDING PROTEIN. [1] ;, mRNA  
sequence.  
ACCESSION AI943364  
VERSION AI943364.1 GI:5708020  
KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprininae; Rasbora; Danio.  
REFERENCE 1 (bases 1 to 43)  
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, V.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
Washu Zebrafish EST Project 1998  
Unpublished (1998)  
On Dec 20, 1995 this sequence version replaced gi:1134662.  
Other\_ESTs: fc79c07.xl  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrfish@watson.wustl.edu  
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:  
www.rzpd.de)  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand

Seq primer: T3 ET from Amersham  
High quality sequence stop: 1.

# FEATURES

source  
Location/Qualifiers  
1. .43  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone\_lib="zebrafish WASHU MPING EST"  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab\_host="Xl1-blue MRF"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; lsl  
strand cDNA was primed with a Not I - oligo(dT)15 primer  
[5']pGACVAGTTCTAGATCGAGCGCGCCCTTTTCTTTTCTTTT3';  
double-stranded cDNA was ligated to Sal I adaptors (BRL),  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BRL). Library was constructed  
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
analysis were selected following oligonucleotide  
hybridization fingerprinting of arrayed clones from  
zebrafish late somitogenesis (26 ss), adult liver or  
embryonic shield stage (5.6 h) libraries. Fingerprint  
data were used to computationally cluster cDNAs, and a  
single cDNA from each cluster was chosen for sequencing.  
In some cases multiple members of the same cluster were  
sequenced to assess clustering parameters or single clones  
were sequenced additional times to assess quality  
control."  
10 a 9 c 11 g 13 t  
BASE COUNT  
ORIGIN  
Query Match 58.0%; Score 11.6; DB 63; Length 43;  
Best Local Similarity 77.8%; Pred. No. 2.8e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 gaacttcctaaaggagg 19  
||||| ||||| |||  
Db 25 GAACGCCCTATGCTAG 42  
RESULT 4  
LOCUS H55376 46 bp DNA EST 07-DEC-1995  
DEFINITION CHR220315 Chromosome 22 exon Homo sapiens genomic clone C22\_397 5',  
mRNA sequence.  
ACCESSION H55376  
VERSION H55376.1 GI:1108242  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 46)  
Trotter, J.A., Long, K.R., Murrell, J.R., Stotler, C.J., Gusella, J.F.  
and Buckler, A.J.  
An expression-independent catalog of genes from human chromosome 22  
Genome Res. 5 (3), 214-224 (1995)  
96159527  
On May 18, 1995 this sequence version replaced gi:811020.  
Contact: Buckler AJ  
Molecular Neurogenetics Unit  
Massachusetts General Hospital  
Building 149, 13th St., Charlestown MA 02129  
Tel: 6177249616  
Fax: 6177265736  
Email: buckler@helix.mgh.harvard.edu  
Insert Length: 588 Std Error: 0.00  
Seq primer: T3  
High quality sequence stop: 356.  
Location/Qualifiers  
1. .46  
FEATURES  
source



analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5,6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 3 a 18 c 16 g 9 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 47; Length 46;  
Best Local Similarity 77.8%; Pred. No. 2 8e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggga 18  
||||| ||||| ||||| |||||  
Db 33 GGAACCCCGGACGGGA 16

RESULT 7  
A1631068  
LOCUS 46 bp mRNA EST 16-DEC-1999  
DEFINITION tz32b08.x1 NCI-CGAP.Ut2 Homo sapiens cDNA clone IMAGE:2290263 3', similar to TR:000599 O00599 CON1.; contains element MER22 repetitive element 1; mRNA sequence.

ACCESSION A1631068  
VERSION A1631068.1 GI:4682398  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 46)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
TITLE Unpublished (1997)  
JOURNAL On Mar 20, 1998 this sequence version replaced gi:2980543.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbrrp/image/image.html

Trace considered overall poor quality  
Insert Length: 1734 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1..46  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2290263"  
/clone\_lib="NCI-CGAP.Ut2"  
/tissue\_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: Salt; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"

BASE COUNT 12 a 17 c 15 g 2 t  
ORIGIN

FEATURES  
source

1..31  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1970259"  
/clone\_lib="Sugano mouse kidney mklia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pME18S-FL3; Site.1: DraIII (CAGCTGTG); Site.2: DraIII (CAGCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CAGCTGTG, 3' site CAGCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTCTCTAAAGCTGCG and 3' end primer CGACCTGAGCTCGAGACA."

BASE COUNT 11 a 7 c 4 g 9 t  
ORIGIN

Query Match 57.0%; Score 11.4; DB 49; Length 46;  
Best Local Similarity 92.3%; Pred. No. 3.6e+04;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ccctaaaggagg 20  
||||| ||||| ||||| |||||  
Db 10 CCTTAAGGGGG 22

RESULT 8  
A1649395

LOCUS 31 bp mRNA EST 30-APR-1999  
DEFINITION uk27h02.x1 Sugano mouse kidney mklia Mus musculus cDNA clone IMAGE:1970259 3' similar to TR:Q64232 Q64232 SC2-SYNAPTIC GLYCOPROTEIN.; mRNA sequence.

ACCESSION A1649395  
VERSION A1649395.1 GI:4730229  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 31)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)

CONTACT: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:986999

High quality sequence stop: 1.  
Location/Qualifiers  
1..31  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1970259"  
/clone\_lib="Sugano mouse kidney mklia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"

Trace considered overall poor quality  
Possible reversed clone: Similarity on wrong strand  
Seq primer: custom primer used  
High quality sequence stop: 1.  
Location/Qualifiers  
1..31  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1970259"  
/clone\_lib="Sugano mouse kidney mklia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"

Trace considered overall poor quality  
Possible reversed clone: Similarity on wrong strand  
Seq primer: custom primer used  
High quality sequence stop: 1.  
Location/Qualifiers  
1..31  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1970259"  
/clone\_lib="Sugano mouse kidney mklia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"

FEATURES  
source

1..31  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1970259"  
/clone\_lib="Sugano mouse kidney mklia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"

Trace considered overall poor quality  
Insert Length: 1734 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1..46  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2290263"  
/clone\_lib="NCI-CGAP.Ut2"  
/tissue\_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: Salt; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"

BASE COUNT 12 a 17 c 15 g 2 t  
ORIGIN

1..46  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2290263"  
/clone\_lib="NCI-CGAP.Ut2"  
/tissue\_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: Salt; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"

BASE COUNT 11 a 7 c 4 g 9 t  
ORIGIN

Query Match 56.0%; Score 11.2; DB 49; Length 31;  
 Best Local Similarity 81.2%; Pred. No. 4.4e+04;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aacttcctctaaaggga 18  
 || ||||| |||||  
 Db 15 AAGTTCCTCCCAAGGCA 30

## RESULT 9

LOCUS A1186043 34 bp mRNA EST 29-OCT-1998  
 DEFINITION qe50f03.x1 Soares fetal lung NbHL19W Homo sapiens cDNA clone  
 IMAGE:1742429 3' similar to SW:RL11\_RAT P25121 60S RIBOSOMAL  
 PROTEIN L11. ; mRNA sequence.

ACCESSION A1186043  
 VERSION A1186043.1 GI:3736681  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 34)  
 AUTHORS Holman, M., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.

TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On May 5, 1995 this sequence version replaced gi:798421.

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Insert Length: 1440 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 1.

## FEATURES

source  
 1..34  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1742429"  
 /clone\_lib="Soares fetal lung NbHL19W"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGGAGCGCGCAATTTTTTTTTTTT-3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT7T3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot - 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo. This library was constructed  
 from the same fetus as the fetal heart library, Soares  
 fetal heart NbHL19W."  
 4 a 6 c 10 g 14 t

BASE COUNT 4 a 6 c 10 g 14 t  
 ORIGIN  
 Query Match 56.0%; Score 11.2; DB 43; Length 34;  
 Best Local Similarity 81.2%; Pred. No. 4.4e+04;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gaacttcctctaaaggg 17  
 || ||||| |||||  
 Db 10 GAAGTTCCTCTATTGGG 25

## RESULT 10

LOCUS H40524 43 bp mRNA EST 16-AUG-1995  
 DEFINITION Y141e11.s1 Soares placenta Nb2HP Homo sapiens cDNA clone  
 IMAGE:141836 3' similar to SP:S42105 S42105 RIBOSOMAL PROTEIN - ;

DEFINITION YP63f01.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone  
 IMAGE:192121 3' similar to SP:RS19\_RAT P17074 40S RIBOSOMAL PROTEIN  
 ; mRNA sequence.

ACCESSION H40524  
 VERSION H40524.1 GI:916576  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 43)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.

TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On May 5, 1995 this sequence version replaced gi:798421.

CONTACT: Willson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 Insert Size: 1124  
 High quality sequence starts: 1  
 High quality sequence stops: 1

Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Insert Length: 1124 Std Error: 0.00  
 Seq primer: Promega -21ml3  
 High quality sequence stop: 1.

High quality sequence stop: 1.  
 Location/Qualifiers  
 1..43

## FEATURES

source  
 1..43  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3761910"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:192121"  
 /clone\_lib="Soares fetal liver spleen lNFLS"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Pac I - oligo(dT) primer  
 [5' AACCTGAAGAATAAATAAGATCTTTTTTTTTTTT 3']  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo."  
 9 a 8 c 18 t 3 others

BASE COUNT 9 a 8 c 18 t 3 others  
 ORIGIN  
 Query Match 56.0%; Score 11.2; DB 23; Length 43;  
 Best Local Similarity 72.2%; Pred. No. 4.6e+04;  
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 aacttcctctaaaggagg 20  
 || ||||| |||||  
 Db 21 AANTCCCCCAATGGGTGG 4

## RESULT 11

LOCUS R70689 49 bp mRNA EST 01-JUN-1995  
 DEFINITION Y141e11.s1 Soares placenta Nb2HP Homo sapiens cDNA clone  
 IMAGE:141836 3' similar to SP:S42105 S42105 RIBOSOMAL PROTEIN - ;



Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouse@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:989941  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: custom primer used  
High quality sequence stop: 1.  
Location/Qualifiers  
source  
1. .40  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1973201"  
/clone\_lib="Sugano mouse kidney mkia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/notes="Organ: kidney; Vector: pME18S-FL3; Site\_1: DraIII  
(CACCATGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested  
and cloned into distinct draIII site of the pME18S-FL3  
vector (5' site CACGTGTG, 3' site CACCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTTCCTGCTCTAAGACGTGCG and 3' end  
primer CGACCTGCAGCTCGACACA."

BASE COUNT 5 a 13 c 12 g 10 t  
ORIGIN  
Query Match 55.0%; Score 11; DB 60; Length 40;  
Best Local Similarity 73.7%; Pred. No. 5.8e+04;  
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 ggaactccctaaaggag 19  
||||| ||||| | |||||  
Db 20 GGAACCTCCCGAGGAG 2

RESULT 14  
AA614588  
LOCUS AA614588 43 bp mRNA EST 06-OCT-1997  
DEFINITION np50601.s1 NCI\_CGAP.Br1.1 Homo sapiens cDNA clone IMAGE:1129705 3' similar to SW:BAT3\_HUMAN P46379 LARGE PROLINE-RICH PROTEIN BAT3 ; mRNA sequence.  
ACCESSION AA614588  
VERSION AA614588.1 GI:2466784  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 43)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 5, 1995 this sequence version replaced gi:797964.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source

1. .43  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1129705"  
/clone\_lib="NCI\_CGAP.Br1.1"  
/sex="female, pooled"  
/tissue\_type="breast"  
/lab\_host="DH10B"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker: 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is not normalized. (The normalized version of this library is NCI\_CGAP.Br2.) Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 11 a 12 c 14 g 6 t  
ORIGIN

Query Match 55.0%; Score 11; DB 36; Length 43;  
Best Local Similarity 73.7%; Pred. No. 5.9e+04;  
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 gaacttcctaaaggagg 20  
||| ||| || |||||  
Db 14 GAAGCTCCAAACGGAGG 32

RESULT 15  
AA411090  
LOCUS AA411090 46 bp mRNA EST 08-AUG-1997  
DEFINITION zt36g09.r1 Soares ovary tumor NBHOT Homo sapiens cDNA clone IMAGE:724480 5' similar to SW:ARND\_HUMAN P41227 N-TERMINAL ACETYLTRANSFERASE COMPLEX ARD1 SUBUNIT HOMOLOG. ; mRNA sequence.  
ACCESSION AA411090  
VERSION AA411090.1 GI:2070227  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 46)  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevasaki, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On Sep 12, 1996 this sequence version replaced gi:1407293.

JOURNAL  
MEDLINE  
COMMENT  
Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Insert length: 953 Std Error: 0.00  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 1.

<b>FEATURES</b>	<b>SOURCE</b>
-----------------	---------------

BASE COUNT	16 a	13 c	12 g	5 t
ORIGIN	M. FALCINO DONADIO.			

Query Match	55.08;	Score 11;	DB 33;	Length 46;
Best Local Similarity	73.78;	Pred. No. 5.9e+04;		
Matches 14;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps 0;				
Qy	2	gaacttcctcaaaaggagg	20	
Db	1	GAACATCGCAATGCGAGG	19	

Search completed: May 23, 2000, 11:00:25  
Job time: 5907 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 11:19:47 ; Search time 68.55 Seconds  
(without alignments)  
37.924 Million cell updates/sec

Title: US-08-945-805-3

Perfect score: 20

Sequence: 1 ggaacttcctaaaggagg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 328916

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
7: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	13.6	68.0	42	4	US-08-343-443B-49
C 2	13.6	68.0	46	4	US-08-381-637-2
C 3	13.2	66.0	45	1	US-08-171-389-130
C 4	13.2	66.0	45	1	US-08-171-389-342
C 5	13.2	66.0	45	2	US-08-123-936-130
C 6	13.2	66.0	45	2	US-08-123-936-342
C 7	13.2	66.0	45	3	US-08-475-228A-130
C 8	13.2	66.0	45	3	US-08-475-228A-342
C 9	13.2	66.0	45	5	US-08-482-080A-130
C 10	13.2	66.0	45	5	US-08-482-080A-342
C 11	13.2	66.0	45	6	PTC-US93-12388-130
C 12	13.2	66.0	45	6	PTC-US93-12388-342
C 13	13.2	66.0	50	1	US-08-171-389-343
C 14	13.2	66.0	50	2	US-08-123-936-343
C 15	13.2	66.0	50	3	US-08-475-228A-343
C 16	13.2	66.0	50	5	US-08-482-080A-343
C 17	13.2	66.0	50	6	PTC-US93-12388-343
C 18	12.4	62.0	39	4	US-08-665-055-1
C 19	12.4	62.0	39	5	US-08-681-297-1
C 20	12.4	62.0	39	5	US-08-342-924-6
C 21	12.2	61.0	22	3	US-08-441-887A-299
C 22	12.2	61.0	22	5	US-08-544-381B-125
C 23	12.2	61.0	23	1	US-08-207-901-21
C 24	12.2	61.0	23	3	US-08-360-051A-28
C 25	12.2	61.0	25	2	US-08-709-733-5
C 26	12.2	61.0	30	3	US-08-360-051A-36
C 27	12.2	61.0	30	3	US-08-360-051A-37

28	12.2	61.0	30	3	US-08-360-051A-38	Sequence 38, Appl
29	12.2	61.0	30	3	US-08-360-051A-39	Sequence 39, Appl
30	12.2	61.0	35	2	US-08-458-423A-9	Sequence 9, Appl
31	12.2	61.0	35	2	US-08-458-424B-9	Sequence 9, Appl
32	12.2	61.0	35	6	PTC-US96-08014-9	Sequence 9, Appl
33	12.2	61.0	44	3	US-08-781-550-12	Sequence 12, Appl
34	12.2	61.0	46	3	US-08-345-861-19	Sequence 19, Appl
35	12.2	61.0	46	3	US-08-479-105A-19	Sequence 19, Appl
36	12.2	61.0	49	3	US-08-441-887A-289	Sequence 289, App
37	12.2	61.0	49	3	US-08-781-550-3	Sequence 3, Appl
38	12.2	61.0	50	3	US-08-781-550-6	Sequence 6, Appl
39	12	60.0	29	1	US-07-977-284A-241	Sequence 241, App
40	12	60.0	29	4	US-08-256-426B-241	Sequence 241, App
C 41	11.8	59.0	38	4	US-08-537-811-29	Sequence 29, Appl
C 42	11.6	58.0	23	4	US-08-778-494B-96	Sequence 96, Appl
C 43	11.6	58.0	25	5	US-09-058-746-15	Sequence 15, Appl
44	11.6	58.0	30	1	US-08-217-210B-15	Sequence 15, Appl
45	11.6	58.0	30	3	US-08-360-051A-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1  
US-08-343-443B-49/c  
; Sequence 49, Application US/08343443B  
; Patent No. 5968734  
; GENERAL INFORMATION:  
; APPLICANT: Aurias, Alain  
; APPLICANT: Delattre, Olivier  
; APPLICANT: Desmaze, Chantal  
; APPLICANT: Melot, Thomas  
; APPLICANT: Peter, Martine  
; APPLICANT: Ploougastel, Beatrice  
; APPLICANT: Thomas, Gilles  
; APPLICANT: Zucman, Jessica  
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
; NUMBER OF SEQUENCES: 129  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Weiser & Associates  
; STREET: 230 South Fifteenth Street  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: AEDIT 1.0 DOS text editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/343,443B  
; FILING DATE: 18-NOV-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR93/00494  
; FILING DATE: 19-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/06123  
; FILING DATE: 20-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weiser, Gerard J.  
; REGISTRATION NUMBER: 19,763  
; REFERENCE/DOCKET NUMBER: 989.6121P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-875-8383  
; TELEFAX: 215-875-8394  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 42 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-343-443B-49

Query Match          68.0%; Score 13.6; DB 4; Length 42;
Best Local Similarity 80.0%; Pred. No. 61;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggagg 20
   ||| ||||| |||||
Db 39 GGAGACTCCCTAAGTGAGG 20

RESULT 2
US-08-381-637-2/c
; Sequence 2, Application US/08381637
; Patent No. 5965124
; GENERAL INFORMATION:
; APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise
; APPLICANT: Weeks-Levy and Patricia Anne Reilly
; TITLE OF INVENTION: Recombinant Vaccines and Method of
; TITLE OF INVENTION: Producing Same
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/381,637
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/986,729
; FILING DATE: 08-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI91-01AA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-381-637-2

Query Match          68.0%; Score 13.6; DB 4; Length 46;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggagg 20
   ||||| ||||| |||||
Db 31 GGAATCCCTCGAGGGAGG 12

RESULT 3
US-08-171-389-130

; LENGTH: 42 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-343-443B-49

Query Match          68.0%; Score 13.6; DB 4; Length 42;
Best Local Similarity 80.0%; Pred. No. 61;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggagg 20
   ||| ||||| |||||
Db 39 GGAGACTCCCTAAGTGAGG 20

RESULT 2
US-08-381-637-2/c
; Sequence 2, Application US/08381637
; Patent No. 5965124
; GENERAL INFORMATION:
; APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise
; APPLICANT: Weeks-Levy and Patricia Anne Reilly
; TITLE OF INVENTION: Recombinant Vaccines and Method of
; TITLE OF INVENTION: Producing Same
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/381,637
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/986,729
; FILING DATE: 08-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI91-01AA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-381-637-2

Query Match          68.0%; Score 13.6; DB 4; Length 46;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggagg 20
   ||||| ||||| |||||
Db 31 GGAATCCCTCGAGGGAGG 12

RESULT 3
US-08-171-389-130

; Sequence 130, Application US/08171389
; Patent No. 5578444
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,389
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human fibrinogen gamma chain gene
; US-08-171-389-130

Query Match          66.0%; Score 13.2; DB 1; Length 45;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggagg 18
   ||| ||| | |||||
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 4
US-08-171-389-342
; Sequence 342, Application US/08171389
; Patent No. 5578444
; GENERAL INFORMATION:
```

APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
APPLICANT: Fry, Kirk E.  
TITLE OF INVENTION: Sequence-Directed DNA Binding  
TITLE OF INVENTION: Molecules, Compositions and Methods  
NUMBER OF SEQUENCES: 641  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,389  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0175/G19P3  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 342:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma  
INDIVIDUAL ISOLATE: chain  
US-08-171-389-342

Query Match 66.0%; Score 13.2; DB 1; Length 45;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggga 18  
||| ||| | ||||| |||  
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 5  
US-08-123-936-130  
Sequence 130, Application US/08123936  
Patent No. 5726014  
GENERAL INFORMATION:  
APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.

APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
TITLE OF INVENTION: Screening Assay for the Detection of  
TITLE OF INVENTION: DNA-Binding Molecules  
NUMBER OF SEQUENCES: 640  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,936  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 130:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human fibrinogen gamma chain gene  
US-08-123-936-130

Query Match 66.0%; Score 13.2; DB 2; Length 45;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggga 18  
||| ||| | ||||| |||  
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 6  
US-08-123-936-342  
Sequence 342, Application US/08123936  
Patent No. 5726014  
GENERAL INFORMATION:  
APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
TITLE OF INVENTION: Screening Assay for the Detection of  
TITLE OF INVENTION: DNA-Binding Molecules  
NUMBER OF SEQUENCES: 640  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA

```
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,936
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 342:
; LENGTH: 45 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma
; US-08-123-936-342

Query Match 66.0%; Score 13.2; DB 2; Length 45;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaacttcctctaaaggga 18
   ||| ||| | |||||
Db 12 GGACCTTACATAAAGGGA 29

RESULT 7
US-08-475-228A-130
; Sequence 130, Application US/08475228A
; Patent No. 5869241
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/475,228A
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 130:
; LENGTH: 45 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human fibrinogen gamma chain gene
; US-08-475-228A-130

Query Match 66.0%; Score 13.2; DB 3; Length 45;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaacttcctctaaaggga 18
   ||| ||| | |||||
Db 12 GGACCTTACATAAAGGGA 29

RESULT 8
US-08-475-228A-342
; Sequence 342, Application US/08475228A
; Patent No. 5869241
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,228A
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
```

; FILING DATE: 17-SEP-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/996,783  
 ; FILING DATE: 23-DEC-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/723,618  
 ; FILING DATE: 27-JUN-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/081,070  
 ; FILING DATE: 22-JUN-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brady, John F.  
 ; REGISTRATION NUMBER: 39,118  
 ; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19p3D1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (650) 324-0880  
 ; TELEFAX: (650) 324-0960  
 ; INFORMATION FOR SEQ ID NO: 130:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 45 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; INDIVIDUAL ISOLATE: Human fibrinogen gamma chain gene  
 ; US-08-482-080A-130

Query Match 66.0%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. NO. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels

Qy 1 ggaacttcctaaagga 18  
||| ||| | |||||  
Db 12 GGAGCTTACATAAGGGA 29

```

RESULT 10
US-08-482-080A-342
; Sequence 342, Application US/08482080A
; Patent No. 6010849
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,080A
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/171,389
; FILING DATE: 20-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993

```

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/996,783  
;; FILING DATE: 23-DEC-1992  
;; PRIOR APPLICATION DATA: US 07/723,618  
;; FILING DATE: 27-JUN-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/081,070  
;; FILING DATE: 22-JUN-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Brady, John F.  
;; REGISTRATION NUMBER: 39,118  
;; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (650) 324-0960  
;; TELEFAX: (650) 324-0960  
;; INFORMATION FOR SEQ ID NO: 342:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 45 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma  
;; INDIVIDUAL ISOLATE: chain  
US-08-482-080A-342

Query Match 66.0%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctcctaaaggga 18  
||| ||| | |||||  
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 11  
PCT-US93-12388-130  
;; Sequence 130, Application PC/TUS9312388  
;; GENERAL INFORMATION:  
;; APPLICANT:  
;; TITLE OF INVENTION: Sequence-Directed DNA Binding  
;; TITLE OF INVENTION: Molecules, Compositions and Methods  
;; NUMBER OF SEQUENCES: 641  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genelabs Technologies, Inc.  
;; STREET: 505 Penobscot Drive  
;; CITY: Redwood City  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94063  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/12388  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/123,936  
;; FILING DATE: 17-SEP-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/996,783  
;; FILING DATE: 23-DEC-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fabian, Gary R.  
;; REGISTRATION NUMBER: 33,875  
;; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0880  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 130:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 45 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; INDIVIDUAL ISOLATE: Human fibrinogen gamma chain gene  
PCT-US93-12388-130

Query Match 66.0%; Score 13.2; DB 6; Length 45;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggaacttcctcctaaaggga 18  
||| ||| | |||||  
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 12  
PCT-US93-12388-342  
;; Sequence 342, Application PC/TUS9312388  
;; GENERAL INFORMATION:  
;; APPLICANT:  
;; TITLE OF INVENTION: Sequence-Directed DNA Binding  
;; TITLE OF INVENTION: Molecules, Compositions and Methods  
;; NUMBER OF SEQUENCES: 641  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genelabs Technologies, Inc.  
;; STREET: 505 Penobscot Drive  
;; CITY: Redwood City  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94063  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/12388  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/123,936  
;; FILING DATE: 17-SEP-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/996,783  
;; FILING DATE: 23-DEC-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fabian, Gary R.  
;; REGISTRATION NUMBER: 33,875  
;; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0880  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 342:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 45 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma  
;; INDIVIDUAL ISOLATE: chain

PCT-US93-12388-342

Query Match 66.0%; Score 13.2; DB 6; Length 45;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaattccctaaaggga 18  
||| ||| | ||| ||| |||  
Db 12 GGAGCTTACATAAAGGGA 29

RESULT 13  
US-08-171-389-343  
; Sequence 343, Application US/08171389  
; Patent No. 5578444  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fty, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; NUMBER OF SEQUENCES: 641  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/171,389  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 343:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma  
; INDIVIDUAL ISOLATE: chain  
US-08-171-389-343

Query Match 66.0%; Score 13.2; DB 1; Length 50;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaattccctaaaggga 18  
||| ||| | ||| ||| |||  
Db 18 GGAGCTTACATAAAGGGA 35

RESULT 14  
US-08-123-936-343  
; Sequence 343, Application US/08123936  
; Patent No. 5726014  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; TITLE OF INVENTION: Screening Assay for the Detection of  
; TITLE OF INVENTION: DNA-Binding Molecules  
; NUMBER OF SEQUENCES: 640  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/123,936  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 343:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma  
; INDIVIDUAL ISOLATE: chain  
US-08-123-936-343

Query Match 66.0%; Score 13.2; DB 2; Length 50;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaattccctaaaggga 18  
||| ||| | ||| ||| |||  
Db 18 GGAGCTTACATAAAGGGA 35

Job time: 5946 sec

RESULT 15  
US-08-475-228A-343  
; Sequence 343, Application US/08475228A  
; Patent No. 5869241  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,228A  
; FILING DATE: 06-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stratford, Carol A.  
; REGISTRATION NUMBER: 34,444  
; REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 343:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Human gene for fibrinogen gamma  
; INDIVIDUAL ISOLATE: chain  
US-08-475-228A-343

Query Match 66.0%; Score 13.2; DB 3; Length 50;  
Best Local Similarity 83.3%; Pred. NO. 1e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 ggaacttcctccctaaaggga 18  
||| ||| | |||||  
DB 18 GGACCTTACATAAGGCA 35

-----

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 10:36:34 ; Search time 236.64 Seconds  
(without alignments)  
-82.217 Million cell updates/sec

Title: US-08-945-805-4  
Perfect score: 20  
Sequence: 1 aacggcatggactgaatcgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 285916

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pl3.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_inl.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13.8	69.0	22	5	AR067584 Sequence
2	12.8	64.0	20	5	A52308 Sequence 12
3	12.8	64.0	20	5	AR050000 Sequence
C 4	12.6	63.0	39	5	A86867 Sequence 7
C 5	12.2	61.0	41	5	A43666 Sequence 10
6	12	60.0	29	5	E15523 Primer. 7/1
7	12	60.0	32	5	A83290 Sequence 6
C 8	12	60.0	49	5	E02141 DNA encodin
9	11.6	58.0	26	5	AR066770 Sequence
10	11.6	58.0	26	5	I71282 Sequence 20
11	11.6	58.0	32	24	E10691 Primer. 10/
C 12	11.6	58.0	35	5	A84223 Sequence 2
C 13	11.6	58.0	35	5	AR064310 Sequence
C 14	11.6	58.0	35	5	I64746 Sequence 42
C 15	11.6	58.0	44	5	AR043085 Sequence
16	11.4	57.0	21	5	AR004650 Sequence
17	11.4	57.0	21	5	AR004652 Sequence
18	11.4	57.0	21	5	AR034562 Sequence
19	11.4	57.0	21	5	AR034564 Sequence
20	11.4	57.0	21	5	I89252 Sequence 13
21	11.4	57.0	21	5	I89254 Sequence 15
22	11.4	57.0	24	5	AR026626 Sequence
23	11.4	57.0	24	5	AR026627 Sequence
24	11.4	57.0	34	5	A56725 Sequence 15
25	11.4	57.0	35	5	A56718 Sequence 8
26	11.4	57.0	36	5	A56733 Sequence 23
27	11.4	57.0	37	5	A56739 Sequence 29
28	11.4	57.0	38	5	A56724 Sequence 14
29	11.4	57.0	39	5	A56717 Sequence 7
30	11.4	57.0	40	5	A56738 Sequence 28
31	11.4	57.0	41	5	A56732 Sequence 22
32	11.4	57.0	42	5	A56731 Sequence 21
33	11.4	57.0	43	5	A56716 Sequence 6
34	11.4	57.0	44	5	A56723 Sequence 13
35	11.4	57.0	45	5	A56722 Sequence 12
36	11.4	57.0	46	5	A56715 Sequence 5
37	11.4	57.0	47	5	A56730 Sequence 20
38	11.4	57.0	48	5	A56737 Sequence 27
C 39	11.2	56.0	20	5	A51149 Sequence 18
C 40	11.2	56.0	20	5	A76974 Sequence 18
C 41	11.2	56.0	21	5	AR067143 Sequence
42	11.2	56.0	24	5	AR050711 Sequence
43	11.2	56.0	30	5	A31714 Mutagenesis
44	11.2	56.0	30	5	I20693 Sequence 6
45	11.2	56.0	35	24	E11202 Probe. 10/1

ALIGNMENTS

```
RESULT 1
AR067584
LOCUS AR067584 22 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 9 from patent US 5851768.
ACCESSION AR067584
VERSION AR067584.1 GI:5998806
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 22)
TITLE de la Chapelle A., Huhtaniemi, I. and Alttomaki, K.
JOURNAL Method for diagnosis of ovarian dysgenesis
PATENT: US 5851768-A 9 22-DEC-1998;
FEATURES
LOCATION/Qualifiers
SOURCE 1..22
BASE COUNT 9 a 4 c 5 g 4 t
ORIGIN

Query Match 69.0%; Score 13.8; DB 5; Length 22;
Best Local Similarity 88.2%; Pred. No. 5.5e+03;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacggcatggactgaat 17
||| ||||| ||||| |||
Db 2 AAGGCAAGGACTGAAT 18

RESULT 2
A52308
LOCUS A52308 20 bp DNA PAT 12-DEC-1997
DEFINITION Sequence 12 from patent EP0721989.
ACCESSION A52308
VERSION A52308.1 GI:2852005
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE
AUTHORS 1 (bases 1 to 20)
TITLE Popoff, M.Y. and Le, G.F.
JOURNAL Oligonucleotides for the detection of salmonella
PATENT: EP 0721989-A 12 17-JUL-1996;
COMMENT PASTEUR INSTITUT (FR)
Other publication JP 8317798 961203
Other publication CA 2167354 960717
Other publication FR 2729392 960719.
FEATURES
LOCATION/Qualifiers
SOURCE 1..20
BASE COUNT 3 a 4 c 7 g 6 t
ORIGIN

Query Match 64.0%; Score 12.8; DB 5; Length 20;
Best Local Similarity 87.5%; Pred. No. 1.9e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acggcatggactgaat 17
||||| ||||| ||||| |
Db 2 ACGCATGGGCTGATT 17

RESULT 3
AR050000
LOCUS AR050000 20 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 15 from patent US 5824795.
ACCESSION AR050000
VERSION AR050000.1 GI:5971992
KEYWORDS
SOURCE
ORGANISM
```

```
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 20)
TITLE Popoff, M.Y. and Fellous, M. LeGuern.
JOURNAL Oligonucleotides for the detection of salmonella
PATENT: US 5824795-A 15 20-OCT-1998;
FEATURES
LOCATION/Qualifiers
SOURCE 1..20
BASE COUNT 3 a 4 c 7 g 6 t
ORIGIN

Query Match 64.0%; Score 12.8; DB 5; Length 20;
Best Local Similarity 87.5%; Pred. No. 1.9e+04;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acggcatggactgaat 17
||||| ||||| ||||| |
Db 2 ACGCATGGGCTGATT 17

RESULT 4
A86867/c
LOCUS A86867 39 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9838317.
ACCESSION A86867
VERSION A86867.1 GI:6735658
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE
AUTHORS 1 (bases 1 to 39)
TITLE Himmlspach, M. and Eibl, J.
JOURNAL FACTOR X ANALOGUES WITH A MODIFIED PROTEASE CLEAVAGE SITE
PATENT: WO 9838317-A 03-SEP-1998;
HIMMELSPACH MICHELE (AT); EIBL JOHANN (AT)
FEATURES
LOCATION/Qualifiers
SOURCE 1..39
BASE COUNT 6 a 13 c 8 g 12 t
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 39;
Best Local Similarity 78.9%; Pred. No. 2.5e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aacggcatggactgaatcg 19
||| ||||| | |||||
Db 27 AACAGCATGCCAGATCG 9

RESULT 5
A43666/c
LOCUS A43666 41 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 10 from Patent WO9507362.
ACCESSION A43666
VERSION A43666.1 GI:2298868
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE
AUTHORS 1 (bases 1 to 41)
TITLE FOX, A. J. and Jones, D. M.
JOURNAL DETECTION AND SPECIATION OF -I(CAMPYLOBACTER)
PATENT: WO 9507362-A 10 16-MAR-1995;
HEALTH LAB SERVICE BOARD (GB)
COMMENT Other publication CA 2168648 950316
Other publication AU 7618394 950327.
FEATURES
LOCATION/Qualifiers
```

```
source
1..41
/organism="unidentified"
/db_xref="taxon:32644"
10 a 7 c 10 g 14 t

BASE COUNT
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 41;
Best Local Similarity 82.4%; Pred. No. 4e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 aacgcgactgactgaat 17
||| ||| ||| ||| |||
Db 36 AACTGCATTGACAGAAAT 20

RESULT 6
E15523
LOCUS E15523 29 bp DNA PAT 28-JUL-1999
DEFINITION Primer.
ACCESSION E15523
VERSION E15523 1 GI:5710206
KEYWORDS JP 1998075787-A/3.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 29)
AUTHORS Houriyoun, K. and Sumiya, J.
TITLE VARIANT TYPE ALANINE AMINOTRANSFERASE AND ITS PRODUCTION
JOURNAL Patent: JP 1998075787-A 24-MAR-1998;
ASahi CHEM IND CO LTD
COMMENT
OC Artificial sequences.
PN JP 1998075787-A/3
PD 24-MAR-1998
PF 02-SEP-1996 JP 1996231540
PI HOURIYOU KAZUO, SUMIYA JUNICHI
PC C12N15/09, C07H21/04, C12N1/21, C12N9/10, C12Q1/48, (C12N1/21, PC
C12R1.19).
PC (C12N9/10, C12R1.19);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
FH key Location/Qualifiers
FH
FT source 1..29
FT /organism='Artificial sequences'.
FEATURES
source
1..29
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
5 a 7 c 13 g 4 t

BASE COUNT
ORIGIN

Query Match 60.0%; Score 12; DB 5; Length 29;
Best Local Similarity 100.0%; Pred. No. 5.1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggcactgactga 15
|||||
Db 17 GGCATGGACTGA 28

RESULT 7
A83290
LOCUS A83290 32 bp DNA PAT 21-JAN-2000
DEFINITION Sequence 6 from Patent WO9850543.
ACCESSION A83290
VERSION A83290.1 GI:6732708
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
```

```
unclassified.
1 (bases 1 to 32)
LeGrand D. and Merot, B.
RECOMBINANT LACTOFERRIN, METHODS OF PRODUCTION FROM PLANTS AND USES
Patent: WO 9850543-A 12-NOV-1998;
LEGRAND DOMINIQUE (FR); BIOCEM S A (FR)
FEATURES
Location/Qualifiers
source
1..32
/organism="unidentified"
/db_xref="taxon:32644"
5 a 7 c 14 g 6 t

BASE COUNT
ORIGIN

Query Match 60.0%; Score 12; DB 5; Length 32;
Best Local Similarity 75.0%; Pred. No. 5.1e+04;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 aacgcgactgactgaatcgg 20
||| ||| ||| ||| |||
Db 8 ACCGGAATGGGCTGGATTGG 27

RESULT 8
E02141/c
LOCUS E02141 49 bp DNA PAT 29-SEP-1997
DEFINITION DNA encoding N terminal of human prokallikrein.
ACCESSION E02141
VERSION E02141.1 GI:2170379
KEYWORDS JP 1989300893-A/1.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 49)
AUTHORS Nakanishi, S.
TITLE PRODUCTION OF HUMAN KALLIKREIN-LIKE PROTEIN
JOURNAL Patent: JP 1989300893-A 1 05-DEC-1989;
NAKANISHI SHIGETADA
COMMENT
OC Artificial gene
OC Artificial sequence; Genes.
OS Human
PN JP 1989300893-A/1
PD 05-DEC-1989
PF 27-MAY-1988 JP 1988129629
PI NAKANISHI SHIGETADA
PC C12N9/64, C12N15/00, (C12N9/64, C12R1.19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=pancreas;
CC *source: clone-phkk-sp;
CC Feature is identified by experimental;
FH key Location/Qualifiers
FH
FT CDS 28..49
FT /product='N-terminal fragment of huamn FT
FT prokallikrein'
FT /partial.
FT Location/Qualifiers
source
1..49
/organism="synthetic construct"
/db_xref="taxon:32630"
4 a 20 c 17 g 8 t

BASE COUNT
ORIGIN

Query Match 60.0%; Score 12; DB 5; Length 49;
Best Local Similarity 100.0%; Pred. No. 5.2e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ggactgaatcgg 20
|||||
```

```
Db 46 GGACTGAATCGG 35

RESULT 9
AR006770 LOCUS 26 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 20 from patent US 5750105.
ACCESSION AR006770
VERSION AR006770.1 GI:3966254
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 26)
AUTHORS Newman,R.A., Hanna,N. and Raab,R.W.
TITLE Recombinant antibodies for human therapy
JOURNAL Patent: US 5750105-A 20 12-MAY-1998;
FEATURES
source 1..26
Location/Qualifiers
BASE COUNT 7 a 6 c 8 g 5 t
ORIGIN

Query Match 58.0%; Score 11.6; DB 5; Length 26;
Best Local Similarity 77.8%; Pred. No. 8.3e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatggactgaatcgg 20
||| ||||| |||
Db 8 CGACATGGACTGGACCTG 25

RESULT 10
I71282 LOCUS 26 bp DNA PAT 03-APR-1998
DEFINITION Sequence 20 from patent US 5681722.
ACCESSION I71282
VERSION I71282.1 GI:3007417
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 26)
AUTHORS Newman,R.A., Hanna,N. and Raab,R.W.
TITLE Recombinant antibodies for human therapy
JOURNAL Patent: US 5681722-A 20 28-OCT-1997;
FEATURES
source 1..26
Location/Qualifiers
BASE COUNT 7 a 6 c 8 g 5 t
ORIGIN

Query Match 58.0%; Score 11.6; DB 5; Length 26;
Best Local Similarity 77.8%; Pred. No. 8.3e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatggactgaatcgg 20
||| ||||| |||
Db 8 CGACATGGACTGGACCTG 25

RESULT 11
E10691 ID E10691 standard; DNA; UNC; 32 BP.
XX AC E10691;
XX AC E10691;
SV E10691.1
XX 08-OCT-1997 (Rel. 52, Created)
DT 08-OCT-1997 (Rel. 52, Last updated, Version 1)
```

---

```
XX DE Primer.
XX KW JP 1996038178-A/14.
XX OS unidentified
OC unclassified.
XX RN [1]
RP 1-32
RA Tanaka S., Niwa H., Tanaka H.:
RT "HUMAN MONOCLONAL ANTIBODY AND PRODUCTION THEREOF, AND PRIMER FOR CLONING
OF GENE THEREOF";
RL Patent number JP 1996038178-A/14, 13-FEB-1996.
RL TANAKA HIDEYUKI, NISSHINBO IND INC.
XX OS None
CC OC Artificial sequences.
CC PN JP 1996038178-A/14
CC PD 13-FEB-1996
CC PF 20-FEB-1995 JP 1995030742
CC PR 18-FEB-1994 JP 94P 21628
CC PI TANAKA SHIGEAKI, NIWA HIROUKI, TANAKA HIDEYUKI
CC PC C12N15/09,C07K16/08,C12N1/21,C12N15/02,C12P21/08,C12Q1/68,
GO1N33/53,
CC PC GO1N33/531,GO1N33/577,(C12N1/21,C12R1:19),(C12P21/08,
C12R1:19);
CC CC Strandedness: Single;
CC CC topology: Linear;
CC CC hypothetical: No;
CC CC anti-sense: No;
CC FH Key
CC FH Location/Qualifiers
CC FT source 1..32
CC FT /organism="Artificial sequences"
XX Key Location/Qualifiers
FH source 1..32
FT /db_xref="taxon:32644"
FT /organism="unidentified"
XX SQ Sequence 32 BP; 5 A; 6 C; 12 G; 6 T; 3 other;

Query Match 58.0%; Score 11.6; DB 24; Length 32;
Best Local Similarity 77.8%; Pred. No. 8.3e+04;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatggactgaatcgg 20
||| ||||| |||
Db 5 CGACATGGACTGGACCTG 22

RESULT 12
A84223/c LOCUS 35 bp DNA PAT 21-JAN-2000
DEFINITION Sequence 2 from Patent WO9846729.
ACCESSION A84223
VERSION A84223.1 GI:6733272
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Murray,J.A. and Lowe,C.R.
TITLE ENZYME ASSAYS
JOURNAL Patent: WO 9846729-A 22-OCT-1998;
MURRAY JAMES AUGUSTUS HENRY (GB); SECR DEFENCE (GB)
FEATURES
source 1..35
Location/Qualifiers
/db_xref="taxon:32644"
```

Search completed: May 23, 2000, 10:36:35  
Job time: 5936 sec

---

Result No.	Query	Score	Match	Length	DB	ID	Description
1	c	13.8	69.0	22	1	T63189	FSH receptor fshr
2		13.2	66.0	50	1	Q34013	Downstream sequenc
3		12.8	64.0	20	1	T37462	Salmonella detecti
4	c	12.6	63.0	39	1	V56757	Human Factor X PCR
5		12.6	63.0	45	1	T70894	Primer WD26 for se
6		12.2	61.0	26	1	V59288	Human telomere rep
7	c	12.2	61.0	41	1	Q86016	Campylobacter prim
8		12	60.0	19	1	V12295	Cyclophilin type P
9		12	60.0	29	1	V16792	PCR primer for hum
10	c	12	60.0	34	1	Q89865	pUC19 637 PCR prim
11		11.8	59.0	23	1	T26944	PCR primer 2F for
12		11.8	59.0	30	1	Q85477	Corynebacteriophag
13	c	11.8	59.0	36	1	V2167	BH3 interacting do
14		11.8	59.0	50	1	V76144	Staphylococcus aur
15		11.6	58.0	26	1	Q35905	Human/monkey heavy
16	c	11.6	58.0	26	1	T62871	Human or monkey VH
17		11.6	58.0	26	1	T92201	Human/monkey heavy
18		11.6	58.0	26	1	T95124	Human or monkey Ig
19	c	11.6	58.0	26	1	V05652	Human/monkey Vh1 e
20		11.6	58.0	26	1	V23762	Primer for Anti-CD
21		11.6	58.0	26	1	V31384	Human or monkey he
22	c	11.6	58.0	30	1	Q89864	VEGF RNA nucleic a
23		11.6	58.0	32	1	Q89870	VEGF RNA nucleic a
24		11.6	58.0	32	1	T18053	Sense primer #1 am
25	c	11.6	58.0	35	1	Q67544	Antisense oligonuc
26		11.6	58.0	35	1	X02224	P. pyralis luc gen
27		11.6	58.0	35	1	X22478	Receptor specific
28	c	11.6	58.0	42	1	Q34016	Downstream sequenc
29		11.6	58.0	50	1	T76979	Staphylococcus aur
30		11.4	57.0	20	1	X00255	Human G3PDH5 contr
31	c	11.4	57.0	20	1	X34942	PCR primer used to
32		11.4	57.0	21	1	Q25814	Clone 45-A primer
33		11.4	57.0	21	1	T31236	Primer GAPDH-Forwa
34	c	11.4	57.0	21	1	V12056	GAPDH (glyceraldeh

```

PI Georges M, Massey JM;
DR WPI; 92-284684/34.
PT Polymorphic bovine DNA markers - used in genetic identification,
PT gene mapping, and selective breeding
PS Table 7; Page 335; 517pp; English.
CC The sequence is that downstream of a bovine microsatellite sequence
CC obtd. by screening a library of bovine MboI DNA fragments of between
CC 250 and 500 bp with an (AC)15 and a (TC)15 oligonucleotide probe.
CC One out of 50 clones cross-hybridised. Assuming independent
CC distribution of microsatellites and MboI sites, the frequency of
CC (76)n >9 microsatellites in the bovine genome is estimated at >100,
CC 000. The sequence information for ca. 230 such bovine microsatellites
CC is summarised in the specification and indexed herein (see below).
CC The sequences upstream and downstream of the microsatellite sequence
CC were used to generate the required PCR primers for in vitro
CC amplification of the corresp. microsatellite (using the program
CC OPTIPRIM). The microsatellites may be used to identify individuals,
CC for parentage testing, and in the genetic mapping of economic trait
CC loci, or genes involved the determinism of economically important
CC traits esp. in cattle, to allow selective breeding.
CC See also Q33501-34437.
SQ Sequence 50 BP; 15 A; 11 C; 13 G; 11 T;

Query Match 66.0%; Score 13.2; DB 1; Length 50;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aacggcgactgactgaatc 18
Db 18 AACCCCATGGACTGAAGC 1

RESULT 3
T37462
ID T37462 standard; DNA; 20 BP.
AC T37462;
DE 25-FEB-1997 (first entry)
DE Salmonella detection primer/probe Iag7.
KW IagA; IagB; Salmonella enterica; primer; probe; HeLa; ss.
OS Synthetic.
PN EP-721989-A1.
PD 17-JUL-1996.
PF 15-JAN-1996; 400098.
PR 16-JAN-1995; FR-000410.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA (INSP ) INST PASTEUR.
PI Le GUERN FELLOUS M, Popoff MY;
DR WPI; 96-322837/33.
PT New nucleic acid of S. enterica ssp. enterica involved in cell
PT invasion - and derived oligo-nucleotide(s) useful as primers and
PT probes for detecting Salmonella in food etc.
PT Claim 2; Page 20; 33pp; French.
PS Oligonucleotides T37451-65 are derived from the newly isolated sequence
CC of the IagA and IagB sequences from Salmonella enterica ssp. enterica
CC serovar typhi. The IagA and IagB sequences (T37466) encode proteins
CC which are involved in the invasion of cultured HeLa cells by S. enterica.
CC This probe can be used to detect bacteria from S. enterica or S. bongori
CC groups I-VI.
SQ Sequence 20 BP; 3 A; 4 C; 7 G; 6 T;

Query Match 64.0%; Score 12.8; DB 1; Length 20;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acggcgactgactgaat 17
Db 2 ACGGCATGGGCTGATT 17

RESULT 4
V56757/c

```

```

ID V56757 standard; DNA; 39 BP.
AC V56757;
DE 27-NOV-1998 (first entry)
DE Human Factor X PCR primer #1005.
KW Factor X; analogue; activation cleavage site; protease; bleeding; human;
KW defect; Factor IX; Factor VII; Factor VIII; haemophilia; gene therapy;
KW PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN W09838317-A1.
PD 03-SEP-1998.
PF 27-FEB-1998; AT0045.
PR 27-FEB-1997; AT-000335.
PA (IMMO ) IMMUNO AG.
PI Darnier F, Eibl J, Fisch A, Himmelspach M, Schlokot U;
DR WPI; 98-481211/41.
PT New factor X analogues with processing site for protease not active
PT on natural protein - and related DNA, is very stable and can be
PT activated in vitro or in vivo without using animal protease(s),
PT particularly for treating disorders of blood coagulation
PS Example 6; Page 39; 86pp; German.
CC V56751-V56775 are primers used in a method resulting in the production
CC of novel human Factor X (F10) analogues. Such analogues have in the
CC region of the natural F10a activation cleavage site, a modification that
CC creates a processing site for a protease that does not naturally cleave
CC F10 in this region. The proteins are used to generate, in vivo or in
CC vitro, F10a analogues that can be used to control bleeding and for
CC treating defects of factors IX, VII or VIII, e.g. in haemophiliacs who
CC have developed antibodies to factors VIII and/or IX. The encoding nucleic
CC acid can be used in gene therapy of the same conditions. The analogues
CC have high stability and can be activated without use of animal enzymes
CC such as trypsin. Only activation is affected, their activity is the same
CC as the natural factor. The analogues can be isolated as a pure
CC single-chain pro-protein (not usually possible because of rapid
CC processing of the native precursor) and this converted to two-chain form
CC by subsequent activation. Activated analogues have good stability and
CC structural integrity and are practically free of inactive intermediates
CC and autolytic decomposition products.
SQ Sequence 39 BP; 6 A; 13 C; 8 G; 12 T;

Query Match 63.0%; Score 12.6; DB 1; Length 39;
Best Local Similarity 78.9%; Pred. No. 4.3e+02;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aacggcgactgactgaatcg 19
Db 27 AACAGCATGACCAGATCG 9

RESULT 5
T70894
ID T70894 standard; DNA; 45 BP.
AC T70894;
DE 29-OCT-1997 (first entry)
DE Primer W026 for secretion of haemoglobin alpha-chain.
KW Primer; PCR; polymerase chain reaction; amplification; haemoglobin;
KW alpha-chain; restriction enzyme; recombinant; beta-chain; plant;
KW oxygen transport; blood; haemorrhage; shock; angioplasty; preservation;
KW organ; transplantation; tumour; sensitisation; gamma-rays;
KW malignant haemopathy; chloroplast; secretion signal; ss.
OS Synthetic.
PN W09704115-A2.
PD 06-FEB-1997.
PF 17-JUL-1996; F01123.
PR 17-JUL-1995; FR-008615.
PA (BIOC-) BIOCEM SA.
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
PI Baudino S, Diercyck W, Gruber V, Lenee P, Marden M;
PI Merot B, Pagnier R, Poyart C;
DR WPI; 97-132653/12.
PT Haem protein prodn. in plant cells contg. DNA encoding protein
PT component - and producing the porphyrin core endogenously, esp. for

```

PT large scale prodn. of virus-free haemoglobin for therapeutic use  
 PS Example V; Page 57; 105pp; French.  
 CC The primers T70893-4 are used to amplify the sequence encoding the  
 CC secretion signal peptide of the sporamine A gene from sweet potato  
 CC tubers. The signal peptide sequence is fused to the first codon of the  
 CC haemoglobin alpha-chain coding sequence. The recombinant construct is  
 CC then used for secretion of the haemoglobin from plant cells. The new  
 CC recombinant haemoglobin molecules are useful where improved oxygen  
 CC transport in the blood is needed, e.g. acute or chronic haemorrhage;  
 CC shock; angioplasty; preservation of organs intended for transplant;  
 CC treatment of solid tumours (sensitisation to gamma-rays) and malignant  
 CC haemopathy.  
 SQ Sequence 45 BP; 11 A; 8 C; 19 G; 7 T;  
 Query Match 63.0%; Score 12.6; DB 1; Length 45;  
 Best Local Similarity 78.9%; Pred. No. 4.3e+02;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 acggcatgactgaatcgg 20  
 |||| |||| |||| ||||  
 Db 22 ACGGAATGGCTGGATTGG 40  
 RESULT 6  
 V59288/c  
 ID V59288 standard; DNA; 26 BP.  
 AC V59288;  
 DT 14-DEC-1998 (first entry)  
 DE Human telomere repeat binding factor primer 2.  
 KW ss; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
 KW telomere; ageing; ataxia telangeiectasia; Down's syndrome; tumour; viral;  
 KW PCR; primer; amplification.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN W09836066-A1.  
 PD 20-AUG-1998.  
 PF 13-FEB-1998; U02765.  
 PR 04-FEB-1998; US-018628.  
 PR 13-FEB-1997; US-800264.  
 PA (U9RQ ) UNIV ROCKEFELLER.  
 PI Bianchi A, De Lange T, Van Steensel B;  
 DT WPI; 98-480769/41.  
 DR Nucleic acid encoding altered telomere repeat binding protein and  
 PT related vectors - transformants, hetero-dimers and antibodies, used  
 PT to inhibit shortening of telomerases caused by ageing or disease,  
 PT also used to extend life of cells in culture  
 PS Example 3; Page 91; 163pp; English.  
 CC The primers V59287-V59289 were used in the production of an altered  
 CC vertebrate telomere repeat binding protein (A-TRF) which has a telomere  
 CC repeat binding factor (TRF) dimerisation domain, and forms a hetero-dimer  
 CC with TRF, preventing it from binding to the specified repeat sequence.  
 CC A-TRF, optionally expressed by gene therapy, is used to inhibit  
 CC shortening of telomeres associated with ageing (for cosmetic purposes)  
 CC and disease, e.g. ataxia telangeiectasia, Down's syndrome, atrophy of the  
 CC skin, age-related macular degeneration, atherosclerosis, tumours and  
 CC viral (including human immune deficiency virus) infection. Cells  
 CC expressing A-TRF also have an increased life span in vitro, e.g. for  
 CC expression of recombinant proteins or where intended for subsequent  
 CC transplant or for testing, eliminating the need for transformation.  
 SQ Sequence 26 BP; 5 A; 7 C; 6 G; 8 T;

Query Match 61.0%; Score 12.2; DB 1; Length 26;  
 Best Local Similarity 82.4%; Pred. No. 6.5e+02;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 ggcgatgactgaatcgg 20  
 ||||| ||||| |||||  
 Db 24 GGCATGAACGTGAACAG 8  
 RESULT 7  
 V16792

Q86016/c  
 ID Q86016 standard; DNA; 41 BP.  
 AC Q86016;  
 DT 20-SEP-1995 (first entry)  
 DE Campylobacter primer Cru0660.  
 KW Speciation; identification; detection; assay; mimic PCR; primer;  
 KW polymerase chain reaction; Campylobacter coli; Campylobacter lari;  
 KW Campylobacter upsaliensis; Campylobacter jejuni; ss.  
 OS Synthetic.  
 PN W09507362-A.  
 PD 16-MAR-1995.  
 PF 09-SEP-1994; G01967.  
 PR 09-SEP-1993; GB-018751.  
 PA (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.  
 PI Fox AJ, Jones DM;  
 DT WPI; 95-123434/16.  
 DR Detection and speciation of campylobacter - by PCR amplification  
 PT of a highly conserved region and restriction endonuclease  
 PT digestion to identify species  
 PS Disclosure; Page 25; 40pp; English.  
 CC Campylobacter speciation into jejuni, coli, upsaliensis or lari,  
 CC or detection of Campylobacter in clinical, environmental or food  
 CC samples is performed by mimic PCR using the primers given in  
 CC Q86015-16; the mimic DNA is given in Q86017.  
 SQ Sequence 41 BP; 10 A; 7 C; 10 G; 14 T;  
 Query Match 61.0%; Score 12.2; DB 1; Length 41;  
 Best Local Similarity 82.4%; Pred. No. 7e+02;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 aacggcatgactgaat 17  
 |||| |||| |||| ||||  
 Db 36 AACTGCATTGACAGAAAT 20  
 RESULT 8  
 V12295  
 ID V12295 standard; DNA; 19 BP.  
 AC V12295;  
 DT 08-JUN-1998 (first entry)  
 DE Cyclophilin type PPIase primer Hppl-5F; -321--303.  
 KW Cyclophilin PPIase; halophilic; archaeobacterium; immunosuppressant;  
 KW cyclosporin A; primer; ss.  
 OS Synthetic.  
 OS Halobacterium cutirubrum.  
 PN J09313184-A.  
 PD 09-DEC-1997.  
 PF 28-MAY-1996; 133353.  
 PR 28-MAY-1996; JP-133353.  
 PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.  
 DR WPI; 98-080075/08.  
 PT New cyclophilin type PPIase gene - purified from halophilic  
 PT archaeobacterium  
 PS Example 5; Page 5; 6pp; Japanese.  
 CC The present sequence represents a primer for cyclophilin type PPIase  
 CC gene from a halophilic archaeobacterium, Halobacterium cutirubrum  
 CC he cyclophilin type PPIase may be combined with an immunosuppressant  
 CC cyclosporin A.  
 SQ Sequence 19 BP; 4 A; 5 C; 6 G; 4 T;  
 Query Match 60.0%; Score 12; DB 1; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 8e+02;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 acggcatgactgact 13  
 ||||| ||||| |||||  
 Db 7 ACGCATGGACT 18  
 RESULT 9  
 V16792

```

ID V16792 standard; DNA; 29 BP.
AC V16792;
DT 24-JUN-1998 (first entry)
DE PCR primer for human alanine aminotransferase.
KW Alanine aminotransferase; mutation; hALT; human; PCR primer;
KW amplify; ss.
OS Synthetic.
OS Homo sapiens.
PN J10075787-A.
PD 24-MAR-1998.
PF 02-SEP-1996; 231540.
PR 02-SEP-1996; JP-231540.
PA (ASAH ) ASAH KASEI KOGYO KK.
DR WPI: 98-244361/22.
PT Recombinant mutant of human alanine aminotransferase - in which at
PT least five N-terminal residues are deleted
PS Disclosure; Page 9; 11pp; Japanese.
CC PCR primers V16791-93 are derived from a known natural human alanine
CC amino transferase (hALT). The specification describes a mutant hALT
CC in which at least 5 amino acids at the N-terminus are deleted. The
CC mutant enzyme retains hALT activity. The mutant hALT can be used
CC to identify individuals carrying the mutant form and to investigate
CC the consequences of carrying this mutation.
SQ Sequence 29 BP; 5 A; 7 C; 13 G; 4 T;

Query Match 60.0%; Score 12; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 ggcattggactga 15
Db 17 GGCATGGACTGA 28

RESULT 10
Q89865
ID Q89865 standard; DNA; 34 BP.
AC Q89865;
DT 02-JAN-1996 (first entry)
DE pUC19 637 PCR primer for generating nucleic acid marker ladder.
KW Nucleic acid marker ladder; DNA; gel electrophoresis; PCR; primer;
KW ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 1
FT /*tag= a
FT /note= "10 nucleotide spacer is present, 5' to
FT thymine at position 1"
FT 1..6
FT misc_feature 1..6
FT /*tag= b
FT /function= NspV_restriction_site
FT 7..14
FT misc_feature 7..14
FT /*tag= c
FT /function= NotI_restriction_site
FT 15..35
FT misc_feature 15..35
FT /*tag= d
FT /note= "Anneals to DNA template"
FT W09511971-A1.
PN 04-MAY-1995.
PD 28-OCT-1994; U12505.
PF 28-OCT-1994; US-142124.
PR (LIFE-) LIFE TECHNOLOGIES INC.
PA Hartley JL;
PI WPI: 95-178864/23.
DR Nucleic acid (NA) marker ladder for estimating mass of NA mol. -
PT comprises at least 3 fragments from complete restriction
PT endonuclease digestion, with fragment lengths being multiples of an
PT integer
PS Example 2; Page 12; 22pp; English.
CC The primers Q89865-Q89870 are PCR primers used in the construction
CC of a new NotI marker ladder. Each primer consists of a spacer sequence,
CC an area of annealing DNA, and one or more restriction sites. The spacing
```

```

CC of the priming sites and the restriction sites is such that when the
CC three PCR products are annealed together to form a circular molecule,
CC the resulting 3500bp plasmids can be cleaved with NotI to give three
CC fragments of 2000, 1000 and 500bp. The marker ladder is useful as a
CC standard for determining the mass of nucleic acid molecules during gel
CC electrophoresis.
SQ Sequence 34 BP; 8 A; 10 C; 11 G; 5 T;

Query Match 60.0%; Score 12; DB 1; Length 34;
Best Local Similarity 75.0%; Pred. No. 8.7e+02;
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 aacggcattggactgaatcgg 20
Db 6 AGCGCCGGTAATGAATCGG 25

RESULT 11
X26944/C
ID X26944 standard; DNA; 23 BP.
AC X26944;
DT 24-JUN-1999 (first entry)
DE PCR primer 2F for amplification of APECED gene fragments.
KW Autoimmune regulator; AIR; immune maturation; immune response;
KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
KW APECED; autoimmune polyglandular syndrome type I; APS I; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN W09915559-A1.
PD 01-APR-1999.
PF 23-SEP-1998; FI0749.
PR 23-SEP-1997; FI-003762.
PA (FIM-) FINNISH IMMUNOTECHNOLOGY LTD..
PI Antonarakis S, Heino M, Krohn K, Kudoh J, Lallioti M,
PI Peterson P, Scott H, Shimizu N;
DR WPI: 99-244390/20.
PT Autoimmune regulator 1 (AIR1) DNA sequence
PS Example 1; Page 12; 59pp; English.
CC PCR primers X26939-47 were used to amplify fragments of the
CC autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
CC (APECED) (also known as autoimmune polyglandular syndrome type I
CC (APS I)) gene. The specification describes autoimmune regulator
CC proteins (AIR-1, AIR-2, and AIR-3). The AIR polypeptides and
CC polynucleotides can be used in methods for the diagnosis and
CC treatment of diseases related to immune maturation and regulation
CC of immune response towards self and nonself. They can be used
CC particularly in the diagnosis and treatment of APECED.
SQ Sequence 23 BP; 6 A; 6 C; 5 G; 6 T;

Query Match 59.0%; Score 11.8; DB 1; Length 23;
Best Local Similarity 86.7%; Pred. No. 1e+03;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 4 ggcattggactgaatc 18
Db 16 GACATGGTCTGAATC 2

RESULT 12
Q85477/C
ID Q85477 standard; DNA; 30 BP.
AC Q85477;
DT 29-SEP-1995 (first entry)
DE Corynebacteriophage beta-toxin 228 fragment A primer #1.
KW Primer; probe; chimeric; vaccinia virus; promoter; hepatitis B virus;
KW epitope; myristoylation site; pre S1 region; immunopotentiating peptide;
KW diphtheria toxin; amplify; E.coli; gpt gene; corynebacteriophage; ss.
OS Synthetic.
PN EP-637631-A.
PD 08-FEB-1995.
PF 29-JUL-1994; 111872.
```

PR 30-JUL-1993; US-099351.  
 PA (IMMO ) IMMUNO AG.  
 PI Dörner F, Falkner F, Pfeleiderer M;  
 DR WPI: 95-068451/10.  
 PT Chimeric vaccinia virus expressing hepatitis B large antigen -  
 PT without the myristoylation site and related antibodies, for  
 PT diagnosis, therapy and vaccines  
 PS Example 1; Column 21; 38pp; German.  
 CC chimeric vaccinia virus containing a promoter functionally coupled to a  
 CC sequence encoding a region of hepatitis B virus including the preS1-B- or  
 CC T-cell epitopes but lacking the myristoylation site of the pre S1 region.  
 CC The virus may also contain a sequence for an immunopotentiating peptide  
 CC such as the CRM228 fragment A (diphtheria toxin cross-reactive material).  
 CC The primers (Q85477-8) were used to amplify the sequence encoding the  
 CC Corynebacteriophage beta-toxin 228 fragment A.  
 SQ Sequence 30 BP; 5 A; 9 C; 7 G; 9 T;

Query Match 59.0%; Score 11.8; DB 1; Length 30;  
 Best Local Similarity 86.7%; Pred. No. 1.1e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 acggcgtgactgaa 16  
 | | | | | | | | | |  
 Db 17 ATGGCATGGGCTGAA 3

RESULT 13  
 V22167  
 ID V22167 standard; cDNA; 36 BP.  
 AC V22167;  
 DT 20-JUL-1998 (first entry)  
 DE BH3 interacting domain death agonist polynucleotide fragment 28.  
 KW Human; BH3 interacting domain death agonist; BID; BCL-2 family;  
 KW apoptosis; regulation; cell death; inflammation; cancer; arthritis;  
 KW autoimmune disease; viral infection; lymphoproliferative; ss.  
 OS Homo sapiens.  
 PN W09809980-A1.  
 PD 12-MAR-1998.  
 PF 09-SEP-1997; U15872.  
 PR 09-SEP-1996; US-706741.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Kormsmeier SJ;  
 DR WPI: 98-193546/17.  
 PT BH3 interacting domain death agonist polypeptide - used for treating  
 PT decreased apoptotic conditions resulting from inflammation etc.  
 PS Disclosure; Page 23; 118pp; English.  
 CC The present sequence represents a BH3 interacting domain death agonist  
 CC (BID) polynucleotide fragment given in the present invention. The  
 CC protein, the DNA encoding it or antisense sequences can be used for  
 CC preventing or treating a decreased apoptotic state of a cell. The  
 CC decreased apoptotic state that is treated results from a disease such as  
 CC cancer, viral infections, lymphoproliferative conditions, arthritis,  
 CC inflammation and autoimmune diseases. Antibodies against the BID protein  
 CC can be used for detecting a BID polypeptide in a cell or population of  
 CC cell. The nucleic acid sequence and the BID protein can also be used for  
 CC treating immunodeficiency disease (including AIDS), senescence,  
 CC neurodegenerative disease, ischaemic and reperfusion cell death,  
 CC infertility and wound-healing. Primers derived from the nucleic acid  
 CC encoding the BID protein can be used for detecting/quantitating the  
 CC protein and for detecting alterations in the nucleic acid encoding the  
 CC BID protein.  
 SQ Sequence 36 BP; 11 A; 5 C; 13 G; 7 T;

Query Match 59.0%; Score 11.8; DB 1; Length 36;  
 Best Local Similarity 86.7%; Pred. No. 1.1e+03;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 acggcgtgactgaa 15  
 | | | | | | | | | |  
 Db 22 AATGGCATGGGCTGAA 36

RESULT 14  
 V76144/c  
 ID V76144 standard; DNA; 50 BP.  
 AC V76144;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #1833.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ss.  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 DR WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1; Page 2072; 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 50 BP; 11 A; 10 C; 9 G; 18 T;

Query Match 59.0%; Score 11.8; DB 1; Length 50;  
 Best Local Similarity 76.5%; Pred. No. 1.2e+03;  
 Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19  
 | | | | | | | | | |  
 Db 48 CNGCATGAATAAATCG 32

RESULT 15  
 Q35905  
 ID Q35905 standard; DNA; 26 BP.  
 AC Q35905;  
 DT 08-JUN-1993 (first entry)  
 DE Human/monkey heavy chain early leader sequence primer VH1.  
 KW Amplify; polymerase chain reaction; heavy; light; chain; variable; Ig;  
 KW PCR; primer; region; lambda; human; constant; antigen; recombinant;  
 KW antibody; chimpanzee; Ag; old world monkey; eczema; immunoglobulin;  
 KW therapeutic; rheumatoid arthritis; ss.  
 OS Synthetic.  
 PN W09302108-A.  
 PD 04-FEB-1993.  
 PF 24-JUL-1992; U06194.  
 PR 25-JUL-1991; US-735064.  
 PR 23-MAR-1992; US-856281.  
 PA (IDEC-) IDEC PHARM CORP.  
 PI Hanna N, Newman RA, Raab RW;

DR WPI; 93-058729/07.  
PT Recombinant antibodies including Old World monkey portion and  
PT human portion - used for treatment of auto-immune diseases,  
PT infectious diseases, AIDS, tumours, diabetes, proliferative  
PT diseases, intestinal inflammations and allergies, etc.  
PS Disclosure; Fig 7aA; 92pp; English.  
CC The sequences given in Q35905-10 represent sense primers which were  
CC used to amplify the human or monkey heavy chain early leader sequence.  
CC These primers bind to the first 23 bases of framework region 1 and  
CC incorporate a SalI site in the amplified product. The amplified  
CC sequences could be used in the production of a recombinant antibody,  
CC comprising a human, chimpanzee or old world monkey immunoglobulin (Ig)  
CC constant region and an antigen (Ag) binding portion of an old world  
CC monkey Ig variable region, where the old world monkeys may be the same  
CC or different. The recombinant antibody may be used as a therapeutic  
CC agent for the treatment of rheumatoid arthritis, eczema and  
CC immunological disorders.  
SQ Sequence 26 BP; 7 A; 6 C; 8 G; 5 T;

Query Match 58.0%; Score 11.6; DB 1; Length 26;  
Best Local Similarity 77.8%; Pred. NO. 1.4e+03;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatggactgaatcgg 20  
||| ||||| |||  
Db 8 CGACATGGACTGGACCTG 25

Search completed: May 23, 2000, 11:21:38  
Job time: 6020 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 11:00:25 ; Search time 1337.41 Seconds  
(without alignments)  
60.613 Million cell updates/sec

Title: US-08-945-805-4  
Perfect score: 20  
Sequence: 1 aacggcatggactgaatcgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues 23194  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*

- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_est45:\*
- 80: gb\_est46:\*
- 81: gb\_est47:\*
- 82: gb\_gss1:\*
- 83: gb\_gss2:\*
- 84: gb\_gss3:\*
- 85: gb\_gss4:\*
- 86: em\_gss1:\*
- 87: em\_gss2:\*
- 88: em\_gss3:\*
- 89: em\_gss4:\*
- 90: gb\_gss5:\*
- 91: gb\_gss6:\*
- 92: gb\_gss7:\*
- 93: gb\_gss8:\*
- 94: gb\_gss9:\*
- 95: em\_gss5:\*
- 96: em\_gss6:\*
- 97: em\_gss7:\*
- 98: em\_gss8:\*
- 99: em\_gss9:\*
- 100: em\_gss10:\*
- 101: em\_gss11:\*
- 102: gb\_gss10:\*
- 103: gb\_gss11:\*
- 104: em\_gss12:\*
- 105: gb\_gss12:\*
- 106: gb\_gss13:\*
- 107: gb\_gss14:\*
- 108: gb\_gss15:\*
- 109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
c 1	12.2	61.0	42	23	H56958	yr09e04.s1
c 2	11.8	59.0	50	23	H55222	CHR220161.C
c 3	11.6	58.0	34	40	AA990297	ua63a02.r
c 4	11.4	57.0	37	23	H40236	yn85h10.r1
c 5	10.8	54.0	45	94	AQ025049	EP(2)1187
c 6	10.8	54.0	49	30	AA228955	ncl4a06.r
c 7	10.6	53.0	20	27	C01261	HUMGS000799
c 8	10.6	53.0	37	40	AA946896	oq53f01.s
c 9	10.6	53.0	40	41	A1014602	cs38c07.s
c 10	10.6	53.0	47	24	H80650	yu77a07.r1
c 11	10.6	53.0	49	42	A1156979	ud08b12.r
c 12	10.4	52.0	33	41	AU007243	AU007243
c 13	10.4	52.0	34	41	AU007259	AU007259
c 14	10.4	52.0	36	38	AA776443	zj50h10.s
c 15	10.4	52.0	37	40	AA918930	ol82a05.s
c 16	10.4	52.0	43	40	AA960099	ub54b07.s
c 17	10.4	52.0	46	43	A1218827	qg92g10.x
c 18	10.4	52.0	46	103	AQ254727	EP(3)3520
c 19	10.4	52.0	49	40	AA930918	vz55d11.r
c 20	10.2	51.0	41	94	AQ025435	EP(X)1392
c 21	10.2	51.0	43	40	AA934191	vz38h02.s
c 22	10.2	51.0	49	25	N51399	yz16h06.s1
c 23	10.2	51.0	49	29	AA133422	z118b12.r
c 24	10	50.0	40	26	W92203	ze15f10.r1
c 25	10	50.0	41	21	T64037	yc05a06.r1
c 26	10	50.0	41	21	T64706	tc25a02.r1
c 27	10	50.0	46	21	T99838	ye71e07.s1
c 28	10	50.0	48	82	HSMC25G04	H.sapiens.D
c 29	10	50.0	49	37	AA673527	vn36f10.r
c 30	10	50.0	49	83	HS4275795	H.sapiens
c 31	9.8	49.0	28	62	AI876227	uj59e02.y
c 32	9.8	49.0	37	41	AI002874	an23d08.s
c 33	9.8	49.0	40	41	AI048812	uc76b03.y
c 34	9.8	49.0	40	83	HS4275800	H.sapiens
c 35	9.8	49.0	45	38	AA809610	nz17e04.s
c 36	9.8	49.0	46	29	AA127131	z187e01.r
c 37	9.8	49.0	46	28	AA764234	vv45h05.r
c 38	9.8	49.0	48	24	H95733	yvl7e08.r1
c 39	9.8	49.0	49	28	AA124369	mq76g11.r
c 40	9.8	49.0	50	40	AI697225	tg17d08.x
c 41	9.6	48.0	28	48	AI583841	tt53e04.x
c 42	9.6	48.0	34	30	AA230484	mw73h07.r
c 43	9.6	48.0	34	37	AA723048	zg83a11.s
c 44	9.6	48.0	34	46	AI433378	tl39c10.x
c 45	9.6	48.0	40	33	AA437989	vd21b08.s

## ALIGNMENTS

RESULT	1
H56958/c	
LOCUS	
DEFINITION	yr09e04.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone IMAGE:204798 3' similar to gb:X00734_cds1 TUBULIN BETA-5 CHAIN (HUMAN); mRNA sequence.
ACCESSION	H56958
VERSION	H56958.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 42) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)

COMMENT	On Apr 14, 1993 this sequence version replaced gi:315287. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu Insert Size: 992 High quality sequence starts: 1 High quality sequence stops: 1 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Trace considered overall poor quality Insert Length: 992 Std Error: 0.00 Seq primer: Promega -21ml3 High quality sequence stop: 1. Location/Qualifiers 1. .42 /organism="Homo sapiens" /db_xref="GDB:3773929" /db_xref="taxon:9606" /clone="IMAGE:204798" /clone_lib="Soares fetal liver spleen lNFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAAATAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT	7 a 14 c 6 g 11 t 4 others
ORIGIN	
Query Match	61.0%; Score 12.2; DB 23; Length 42;
Best Local Similarity	73.7%; Pred. No. 1.1e+04;
Matches	14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy	1 aacggcgatggactgaatcg 19 
Db	25 AAGGCGATGGACNANATGG 7
RESULT	2
H55222	
LOCUS	
DEFINITION	H55222 50 bp DNA EST 07-DEC-1995 CHR220161 Chromosome 22 exon Homo sapiens genomic clone C22_203 5', mRNA sequence.
ACCESSION	H55222
VERSION	H55222.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 50) Trofatter,J.A., Long,K.R., Murrell,J.R., Stotler,C.J., Gusella,J.F. and Buckler,A.J.
TITLE	An expression-independent catalog of genes from human chromosome 22
JOURNAL	Genome Res. 5 (3), 214-224 (1995)
MEDLINE	96159527
COMMENT	On May 8, 1995 this sequence version replaced gi:801194. Contact: Buckler AJ Massachusetts General Hospital Building 149, 13th St., Charlestown MA 02129 Tel: 6177249616

Fax: 6177265736  
Email: buckler@helix.mgh.harvard.edu  
Insert Length: 916 Std Error: 0.00  
Seq primer: T3  
High quality sequence stop: 269.  
Location/Qualifiers

## FEATURES

source

1. .50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="C22.203"  
/clone\_lib="Chromosome 22 exon"  
/lab\_host="E. coli DH5a"  
/note="Vector: pBluescriptIIKS+; Site.1: Sal I; Site.2: Bam HI (destroyed); Exons were isolated from human chromosome 22 specific cosmids using a modification of the method of exon amplification (Proc. Natl. Acad. Sci. USA 88:4005-4009, 1991). Amplified exons were digested with Sal I and Bgl II and subsequently cloned into pBluescriptIIKS+ at the Sal I and Bam HI sites."  
9 a 16 c 16 g 9 t

BASE COUNT  
ORIGIN

Query Match 59.0%; Score 11.8; DB 23; Length 50;  
Best Local Similarity 86.7%; Pred. No. 1.8e+04;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 gcattgactgaatcg 19  
|||||  
Db 17 GCATGGACTGTGTCG 31

RESULT 3

AA990297

LOCUS

AA990297 34 bp mRNA EST 02-JUN-1998  
ua63a02.rl Soares\_thymus\_2NdbMT Mus musculus cDNA clone  
IMAGE:1362122 5' similar to TR:014684 O14684 FIG12. [1] ;, mRNA  
sequence.

ACCESSION

AA990297

VERSION

AA990297.1 GI:3175661

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE

The WashU-HMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

On Jan 19, 1998 this sequence version replaced gi:2285181.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:895342  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES

source

1. .34  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

/clone="IMAGE:1362122"  
/clone\_lib="Soares\_thymus\_2NdbMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCCGGTTTTTTTTTTTTTTTTTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through two  
rounds of normalization, and was constructed by Bento  
Soares and M.Fatima Bonaldo."  
10 a 12 c 5 g 7 t

BASE COUNT  
ORIGIN

Query Match 58.0%; Score 11.6; DB 40; Length 34;  
Best Local Similarity 77.8%; Pred. No. 2.1e+04;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 aacggcatggactgaatc 18  
|||||  
Db 8 AACGACATGGAGACAATC 25

RESULT 4

H40236/c

LOCUS

H40236

DEFINITION

H40236

ACCESSION

H40236

VERSION

H40236.1 GI:916288

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

On May 5, 1995 this sequence version replaced gi:797696.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1506  
High quality sequence starts: 1  
High quality sequence stops: 1  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Trace considered overall poor quality  
Insert Length: 1506 Std Error: 0.00  
Seq primer: M13RPI  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES

source

1. .37  
/organism="Homo sapiens"  
/db\_xref="GDB:3837302"  
/db\_xref="taxon:9606"  
/clone="IMAGE:175267"

```

/clone_lib="Soares adult brain N2b5HB55y"
/sex="Male"
/dev_stage="55-year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: brain; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTTTTTTTTTTTT 3'], the
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 53. Library constructed by Bento
Soares and M.Fatima Bonaldo. The adult brain RNA was
provided by Dr. Donald H. Gilden. Tissue was acquired
17-18 hours after death which occurred in consequence of a
ruptured aortic aneurysm. RNA was prepared from a pool of
tissues representing the following areas of the brain:
frontal, parietal, temporal and occipital cortex from the
left and right hemispheres, subcortical white matter,
basal ganglia, thalamus, cerebellum, midbrain, pons and
medulla."
BASE COUNT      9 a      14 c      6 g      5 t      3 others
ORIGIN

Query Match      57.0%; Score 11.4; DB 23; Length 37;
Best Local Similarity 92.3%; Pred. No. 2.8e+04;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2  acgcgatggactg 14
      | | | | | | | | | |
Db  27 ATGCGATGGACTG 15

RESULT  5
AQ025049
LOCUS      45 bp      DNA      GSS      14-OCT-1998
DEFINITION EP(2)1187 Drosophila melanogaster EP line Drosophila melanogaster
genomic Sequence recovered from 5' end of P element, genomic survey
sequence.
ACCESSION  AQ025049
VERSION     AQ025049.1 GI:3265401
KEYWORDS   GSS.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 45)
AUTHORS   Rehm,E.J. and Rubin,G.M.
TITLE     The BDP gene disruption project: single EP element insertions
JOURNAL   Unpublished (1998)
COMMENT   Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu

Sequence recovery method was inverse PCR.

Sequence orientation is forward strand relative to 5' end of P
element

The P element insertion position is base 38 in the 45 bases. This
insertion position refers to the first base of the 8 base target
recognition sequence.

Class: transposon-tagged.
Location/Qualifiers
source      1. .45
            /organism="Drosophila melanogaster"

```

```

/db_xref="taxon:7227"
/clone_lib="Drosophila melanogaster EP line"
/notes="Inverse PCR was performed on Drosophila
melanogaster strains each of which contains a single EP
transposable element insertion. (The generation of these
insertion strains is described in Rorth P, Szabo K, Bailey
A, Laverty T, Rehm J, Rubin GW, Weigmann K, Milan M, Benes
V, Ansorge W, Cohen SM. 1998. Systematic gain-of-function
genetics in Drosophila. Development 6:1049-1057.) The
resultant fragment for each strain was directly sequenced
to determine the genomic sequence at the site of
insertion. Details of the protocols used can be found at
http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT      13 a      13 c      13 g      6 t
ORIGIN

Query Match      54.0%; Score 10.8; DB 94; Length 45;
Best Local Similarity 85.7%; Pred. No. 6e+04;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  1  aacggcatggactg 14
      | | | | | | | | | |
Db  16 AACGGCTTGACTG 29

RESULT  6
AA228955
LOCUS      49 bp      mRNA      EST      20-AUG-1997
DEFINITION nc14a06.r1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1008082
similar to TR:G929632 G929632 PUTATIVE ORF. ; mRNA sequence.
ACCESSION  AA228955
VERSION     AA228955.1 GI:1875049
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 49)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Mar 10, 1997 this sequence version replaced gi:1851774.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

```

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

source

1. .49

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1008082"

/clone\_lib="NCI\_CGAP\_Prl"

/sex="Male"

/dev\_stage="45 years old"

/lab\_host="DH10B"

/notes="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st

strand cDNA was primed with oligo(dT)17 on 50 ng of

DNase-treated, total cellular RNA obtained from

5,000-10,000 microdissected, histologically normal

prostate epithelial cells. Double-stranded cDNA was

ligated to EcoRI adaptors, 5 cycles of PCR applied to the

cdna with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UBC-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT	13 a	7 c	13 g	16 t
ORIGIN	Conserved 57% and 44% from human.			
Query Match	54.0%; Score 10.8; DB 30; Length 49;			
Best Local Similarity	85.7%; Pred. No. 6.1e+04;			
Matches 12: Conservative	0; Mismatches 2; Indels 0; Gaps 0;			

RESULT	7
C01261/c	
LOCUS	C01261 20 bp mRNA EST 23-JUL-1996
DEFINITION	HUMGS0007998 Human adult (K.Okubo) Homo sapiens cDNA, mRNA sequence.

REFERENCE 1 (bases 1 to 20)

AUTHORS	OKUBO, K.
TITLE	BodyMap; human gene expression database
JOURNAL	Unpublished (1995)
COMMENT	On May 18, 1995 this sequence version replaced qi:811063.

On May 18, 1995, this sequence version replaced g1:811063.  
Contact: Okubo.K.  
Institute for Molecular and Cellular Biol  
Osaka University  
1-3, Yamada-oka, Suita, Osaka Pref. 565, Japan  
Tel.: 06-877-5111(ex.3315)  
Email: kousaku@imcb.osaka-u.ac.jp  
Human Gene Signature: 3'-directed cDNA sequence. We are not  
submitting the same cDNA sequence redundantly to DDBJ since 1993.  
for the abundance information of clones with this sequence in this  
library and as well as in other 3'-directed libraries, see '  
<http://www.imcb.osaka-u.ac.jp/bodymap/>'. The sequences of the clones  
represented by this GS sequences is also found there.

```

FEATURES
  source
    1. .20
      Location/Qualifiers
        representation="text"
        db_xref="taxon:9606"
        clone_lib="Human adult (K. Okubo)"
        dev_stage="adult"

```

Query Match	53.0%	Score	10.6;	DB	27;	Length	20;
Best Local Similarity	76.5%	Pred. No.	6.4e+04;				
Matches	13;	Conservative	0;	Mismatches	4;	Indels	0;
						Gaps	0;

RESULT	8
AA946896	
LOCUS	
DEFINITION	37 bp mRNA EST
	AA946896
	Oq53f01.s1 NCI_CGAP_kid5 Homo sapiens cDNA clone IMAGE:1590073 3'
	similar to TR:O07708 O07708 HYPOTHETICAL 23.7 KD PROTEIN. ; mRNA
	sequence.
ACCESSION	AA946896
	23-JUL-1998

VERSION AA946896.1 GI:3110291  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

ORGANISM *homo sapiens*  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 37)

NCI-CCGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2152699.
Contact: Robert Strausberg, Ph.D.

Contact: Robert Strausberg, F.M.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can  
be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 1105 Std Error: 0.00  
Seq primer: -40ml3 fwd. Et from Amersham  
High quality sequence stop: 1.

```

BASE COUNT      10 a      10 c      4 g      13 t
ORIGIN
                                constructed by Bento Soares and M. Patricia Bonatto.

Query Match      53.0%      score 10.6;      DB 40;      Length 37;
Best Local Similarity 76.5%      pred. No. 7.3e+04;
Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0

```

RESULT	9
AI014602	
LOCUS	AI014602 40 bp mRNA EST 27-JUL-1998
DEFINITION	OS38C07.s1 NCI_CGAP-BR-2 Homo sapiens cDNA clone IMAGE:1607628 3' similar to TR:Q61123 Q61123 MATERNAL EMBRYONIC MESSAGE 3 ; , mRNA sequence

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 40)  
AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute. Cancer Genome Anatomy Project. (CGAP)

**JOURNAL**  
**COMMENT**  
 Tumor Gene Index  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1397854.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
 Insert Length: 931 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 1.

**FEATURES**  
 source  
 1..40  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NCI-CGAP\_Br2"  
 /sex="female, pooled"  
 /tissue\_type="breast"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. This library is the normalized version of NCI-CGAP\_Br1.1. Library was constructed by Bento Soares and M. Fatima Bonaudo."

**BASE COUNT** 9 a 9 c 8 g 14 t  
**ORIGIN**

Query Match 53.0%; Score 10.6; DB 41; Length 40;  
 Best Local Similarity 76.5%; Pred. No. 7.4e+04;  
 Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcattggaactgcg 19  
 |||||  
 Db 2 CGTCATGCACTGATTG 18

**RESULT 10**  
**H80650/c**  
**LOCUS**  
**DEFINITION** H80650 47 bp mRNA EST 09-NOV-1995  
 yu77a07.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
 IMAGE:239796 5' similar to gb:M22324 AMINOPEPTIDASE N (HUMAN);,  
 mRNA sequence.

**ACCESSION** H80650  
**VERSION** H80650.1 GI:1058739  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
**REFERENCE**  
**AUTHORS** 1 (bases 1 to 47)  
 Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasaki, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
**TITLE** The WashU-Merck EST Project  
**JOURNAL** Unpublished (1995)  
**COMMENT** On Sep 21, 1992 this sequence version replaced gi:279126.  
 Contact: Wilson RK

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewartson.wustl.edu  
 High quality sequence starts: 1  
 High quality sequence stops: 1  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Insert Length: 1436 Std Error: 0.00  
 Seq primer: M13RP1  
 High quality sequence stop: 1.

**FEATURES**  
 Location/Qualifiers  
 1..47  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3788749"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:239796"  
 /clone\_lib="Soares fetal liver spleen INFLS"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
 with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;  
 1st strand cDNA was primed with a Pac I - oligo(dT) primer  
 [5' ACTGCGAGATTAATTAAGATCTTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M. Fatima Bonaudo."

**BASE COUNT** 12 a 14 c 12 g 9 t  
**ORIGIN**

Query Match 53.0%; Score 10.6; DB 24; Length 47;  
 Best Local Similarity 76.5%; Pred. No. 7.7e+04;  
 Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acggcattggaactgc 18  
 |||||  
 Db 39 AGGCGATGCTACTGCTC 23

**RESULT 11**  
**A1156979/c**  
**LOCUS**  
**DEFINITION** A1156979 49 bp mRNA EST 30-SEP-1998  
 ud08h12.r1 Soares\_NMPu Mus musculus cDNA clone IMAGE:1434599 5',  
 similar to SW:RL39\_HUMAN P02404 60S RIBOSOMAL PROTEIN L39.;, mRNA  
 sequence.

**ACCESSION** A1156979  
**VERSION** A1156979.1 GI:3685448  
**KEYWORDS** EST.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE**  
**AUTHORS** 1 (bases 1 to 49)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubouque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wyllie, I., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
**TITLE** The WashU-HHMI Mouse EST Project  
**JOURNAL** Unpublished (1996)  
**COMMENT** On Jan 19, 1998 this sequence version replaced gi:2151042.  
 Contact: Maria M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800

FEATURES	SOURCE
1. <b>Feature 1</b>	Source 1
2. <b>Feature 2</b>	Source 2
3. <b>Feature 3</b>	Source 3
4. <b>Feature 4</b>	Source 4
5. <b>Feature 5</b>	Source 5
6. <b>Feature 6</b>	Source 6
7. <b>Feature 7</b>	Source 7
8. <b>Feature 8</b>	Source 8
9. <b>Feature 9</b>	Source 9
10. <b>Feature 10</b>	Source 10
11. <b>Feature 11</b>	Source 11
12. <b>Feature 12</b>	Source 12
13. <b>Feature 13</b>	Source 13
14. <b>Feature 14</b>	Source 14
15. <b>Feature 15</b>	Source 15
16. <b>Feature 16</b>	Source 16
17. <b>Feature 17</b>	Source 17
18. <b>Feature 18</b>	Source 18
19. <b>Feature 19</b>	Source 19
20. <b>Feature 20</b>	Source 20
21. <b>Feature 21</b>	Source 21
22. <b>Feature 22</b>	Source 22
23. <b>Feature 23</b>	Source 23
24. <b>Feature 24</b>	Source 24
25. <b>Feature 25</b>	Source 25
26. <b>Feature 26</b>	Source 26
27. <b>Feature 27</b>	Source 27
28. <b>Feature 28</b>	Source 28
29. <b>Feature 29</b>	Source 29
30. <b>Feature 30</b>	Source 30
31. <b>Feature 31</b>	Source 31
32. <b>Feature 32</b>	Source 32
33. <b>Feature 33</b>	Source 33
34. <b>Feature 34</b>	Source 34
35. <b>Feature 35</b>	Source 35
36. <b>Feature 36</b>	Source 36
37. <b>Feature 37</b>	Source 37
38. <b>Feature 38</b>	Source 38
39. <b>Feature 39</b>	Source 39
40. <b>Feature 40</b>	Source 40
41. <b>Feature 41</b>	Source 41
42. <b>Feature 42</b>	Source 42
43. <b>Feature 43</b>	Source 43
44. <b>Feature 44</b>	Source 44
45. <b>Feature 45</b>	Source 45
46. <b>Feature 46</b>	Source 46
47. <b>Feature 47</b>	Source 47
48. <b>Feature 48</b>	Source 48
49. <b>Feature 49</b>	Source 49
50. <b>Feature 50</b>	Source 50
51. <b>Feature 51</b>	Source 51
52. <b>Feature 52</b>	Source 52
53. <b>Feature 53</b>	Source 53
54. <b>Feature 54</b>	Source 54
55. <b>Feature 55</b>	Source 55
56. <b>Feature 56</b>	Source 56
57. <b>Feature 57</b>	Source 57
58. <b>Feature 58</b>	Source 58
59. <b>Feature 59</b>	Source 59
60. <b>Feature 60</b>	Source 60
61. <b>Feature 61</b>	Source 61
62. <b>Feature 62</b>	Source 62
63. <b>Feature 63</b>	Source 63
64. <b>Feature 64</b>	Source 64
65. <b>Feature 65</b>	Source 65
66. <b>Feature 66</b>	Source 66
67. <b>Feature 67</b>	Source 67
68. <b>Feature 68</b>	Source 68
69. <b>Feature 69</b>	Source 69
70. <b>Feature 70</b>	Source 70
71. <b>Feature 71</b>	Source 71
72. <b>Feature 72</b>	Source 72
73. <b>Feature 73</b>	Source 73
74. <b>Feature 74</b>	Source 74
75. <b>Feature 75</b>	Source 75
76. <b>Feature 76</b>	Source 76
77. <b>Feature 77</b>	Source 77
78. <b>Feature 78</b>	Source 78
79. <b>Feature 79</b>	Source 79
80. <b>Feature 80</b>	Source 80
81. <b>Feature 81</b>	Source 81
82. <b>Feature 82</b>	Source 82
83. <b>Feature 83</b>	Source 83
84. <b>Feature 84</b>	Source 84
85. <b>Feature 85</b>	Source 85
86. <b>Feature 86</b>	Source 86
87. <b>Feature 87</b>	Source 87
88. <b>Feature 88</b>	Source 88
89. <b>Feature 89</b>	Source 89
90. <b>Feature 90</b>	Source 90
91. <b>Feature 91</b>	Source 91
92. <b>Feature 92</b>	Source 92
93. <b>Feature 93</b>	Source 93
94. <b>Feature 94</b>	Source 94
95. <b>Feature 95</b>	Source 95
96. <b>Feature 96</b>	Source 96
97. <b>Feature 97</b>	Source 97
98. <b>Feature 98</b>	Source 98
99. <b>Feature 99</b>	Source 99
100. <b>Feature 100</b>	Source 100

Query Match	53.0%;	Score 10.6;	DB 42;	Length 49;
Best Local Similarity	76.5%;	Pred. No. 7.8e+04;		
Matches 13; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

RESULT	12
AU007243	
LOCUS	33 bp mRNA EST
DEFINITION	AU007243 Schizosaccharomyces pombe late log phase cDNA
	Schizosaccharomyces pombe cDNA clone spc01664, mRNA sequence.
ACCESSION	AU007243
VERSION	AU007243.1 GI:3343701
KEYWORDS	EST.

```

FEATURES
source
1. .33
Location/Qualifiers
/organism="Schizosaccharomyces pombe"
/strain="p72"
/db_xref="taxon:4896"
/clone_lib="spc01664"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mpl9; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mpl9 DNA and the direction of DNA

```

Query Match 52.0%; Score 10.4; DB 41; Length 33;  
Best Local Similarity 91.7%; Pred. No. 9.1e+04;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT	13				
AU007259					
LOCUS		34 bp	mRNA	EST	31-JUL-1998
DEFINITION	AU007259 Schizosaccharomyces pombe late log phase cDNA				
	Schizosaccharomyces pombe cDNA clone spc01683, mRNA sequence.				
ACCESSION	AU007259				
VERSION	AU007259.1	GI:3343717			
KEYWORDS	EST.				

FEATURES	Location/Qualifiers
source	1. :34

```

/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc01683"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mpl9; The cDNA library of Schizosaccharomyces pombe was prepared by cloning cDNA into the SmaI site of M13mpl9 DNA and the direction of DNA sequences was not always from 5' to 3'. The cDNA data of Schizosaccharomyces pombe are available for searching on the World Wide Web. (URL, http://www.nirs.go.jp)"
12 a 3 c 10 a 7 t 2 others
BASE COUNT

```

Query Match 52.0%; Score 10.4; DB 41; Length 34;  
Best Local Similarity 91.7%; Pred. No. 9.1e+04;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 gcatggactgaa 16  
|||  
Db 10 GCTTGGACTGAA 21

RESULT	14
AA776443/C	
LOCUS	AA776443    36 bp     mRNA                EST           05-FEB-1998
DEFINITION	cj50hi0.s1 Soares fetal_liver_spleen_INFLS.S1 Homo sapiens cdna clone IMAGE:453763 3' similar to gb L28107 TRRRRH Trichoderma reesei 25S ribosomal (rRNA); mRNA sequence.
ACCESSION	AA776443

AA776443.1 GI:2835777  
EST.  
SOURCE  
ORGANISM  
human.

REFERENCE  
AUTHORS  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marr,A.M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
Unpublished (1997)

TITLE  
JOURNAL  
COMMENT  
On Jan 17, 1998 this sequence version replaced gi:2044737.  
Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.

FEATURES  
Location/Qualifiers  
1..36  
    /organism="Homo sapiens"  
    /db\_xref="GDB:1390119"  
    /db\_xref="taxon:9606"  
    /clone="IMAGE:453763"  
    /clone\_lib="Soares\_fetal\_liver\_spleen\_INFLS\_SI"  
    /sex="male"  
    /dev\_stage="20 week-post conception fetus"  
    /lab\_host="DH10B (ampicillin resistant)"  
    /note="Organ: Liver and Spleen; Vector: pRT73D (Pharmacia)  
This is a modified polylinker; Site\_1: Pac I and cloned into the Pac I  
with a substituted polylinker; Site\_2: Eco RI;  
This is a subtracted version of the original Soares fetal  
liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I - oligo(3T) primer [5',  
AAGCTGAAGATTAATTAGACTCTTTTGTGGTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pRT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT  
ORIGIN  
3 a 16 c 6 g 11 t

Query Match 52.0%; Score 10.4; DB 38; Length 36;  
Best Local Similarity 70.0%; Pred. No. 9.2e+04;  
Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 aacggcatggaactgaatcgg 20  
||||| | | | | | | | | |  
Db 34 AACGGCGCTTGCGGAATCAG 15

RESULT 15  
AA918930  
LOCUS  
DEFINITION  
AA918930 37 bp mRNA EST 17-APR-1998  
similar to SW:COX2\_HUMAN P00403 CYTOCHROME C OXIDASE POLYPEPTIDE II  
;; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 37)

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 11:19:49 ; Search time 68.55 Seconds  
(without alignments)  
37,924 Million cell updates/sec

Title: US-08-945-805-4  
Perfect score: 20  
Sequence: 1 aacggcatgactgaatcgg 20  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 328916

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	13.8	69.0	22	3	US-08-531-070A-9
2	12.8	64.0	20	2	US-08-586-272-15
3	12.2	61.0	26	4	US-09-018-628-20
4	12.2	61.0	26	5	US-09-273-378-20
5	12.2	61.0	41	5	US-08-604-991-10
6	11.8	59.0	36	4	US-08-924-635A-28
7	11.6	58.0	26	1	US-08-478-039-20
8	11.6	58.0	26	2	US-08-476-349A-20
9	11.6	58.0	26	4	US-08-859-998-657
10	11.6	58.0	32	4	US-08-859-998-275
11	11.6	58.0	35	1	US-08-451-240-42
12	11.6	58.0	35	3	US-08-470-846A-38
13	11.6	58.0	35	6	PCT-US94-12591-42
14	11.6	58.0	21	2	US-08-096-762-209
15	11.4	57.0	21	1	US-08-041-538-13
16	11.4	57.0	21	1	US-08-041-538-15
17	11.4	57.0	21	2	US-08-463-642-13
18	11.4	57.0	21	2	US-08-463-642-15
19	11.4	57.0	21	2	US-08-455-602-13
20	11.4	57.0	21	2	US-08-455-602-15
21	11.4	57.0	21	3	US-08-465-157-13
22	11.4	57.0	21	3	US-08-465-157-15
23	11.4	57.0	21	6	PCT-US91-09422-13
24	11.4	57.0	21	6	PCT-US91-09422-15
25	11.4	57.0	22	4	US-08-882-435-4
26	11.4	57.0	24	3	US-08-203-532F-18
27	11.4	57.0	24	3	US-08-203-532F-19

c	28	11.2	56.0	21	3	US-08-117-952-491	Sequence 491, Appl
	29	11.2	56.0	24	1	US-07-841-662-18	Sequence 18, Appl
	30	11.2	56.0	24	1	US-08-209-797-18	Sequence 18, Appl
	31	11.2	56.0	24	2	US-08-669-685-18	Sequence 18, Appl
	32	11.2	56.0	24	5	US-09-103-486-18	Sequence 18, Appl
	33	11.2	56.0	24	6	PCT-US93-01557-18	Sequence 18, Appl
	34	11.2	56.0	28	4	US-08-859-998-1292	Sequence 1292, Ap
c	35	11.2	56.0	28	6	PCT-US95-15716-90	Sequence 90, Appl
	36	11.2	56.0	30	1	US-08-150-331-6	Sequence 6, Appl
c	37	11.2	56.0	31	4	US-08-859-998-258	Sequence 258, App
	38	11.2	56.0	36	1	US-08-291-932A-473	Sequence 473, App
	39	11.2	56.0	36	1	US-08-291-932A-662	Sequence 662, App
	40	11.2	56.0	36	2	US-08-682-218-28	Sequence 28, Appl
	41	11.2	56.0	37	2	US-08-682-218-27	Sequence 27, Appl
	42	11.2	56.0	39	1	US-08-456-840-3	Sequence 3, Appl
	43	11.2	56.0	39	1	US-08-456-840-5	Sequence 5, Appl
	44	11.2	56.0	39	2	US-08-266-407A-3	Sequence 3, Appl
	45	11.2	56.0	39	2	US-08-266-407A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1  
US-08-531-070A-9  
; Sequence 9, Application US/08531070A  
; Patent No. 5851768  
; GENERAL INFORMATION:  
; APPLICANT: de la Chapelle, Albert  
; APPLICANT: Aittomaki, Kristiina  
; APPLICANT: Huhtaniemi, Ilpo  
; TITLE OF INVENTION: Method For Diagnosis Of Ovarian Dysgenesis  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,070A  
; FILING DATE: 20-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gass, David A.  
; REGISTRATION NUMBER: 38,153  
; REFERENCE/DOCKET NUMBER: 28113/32879  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-531-070A-9

Query Match 69.0%; Score 13.8; DB 3; Length 22;  
Best Local Similarity 88.2%; Pred. No. 59;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 aacggcatgactgaat 17  
|| |||| |||||||||

Db 2 AAAGCAAGGACTGAAT 18

RESULT 2

US-08-586-272-15

Sequence 15, Application US/08586272

Patent No. 5824795

GENERAL INFORMATION:

APPLICANT: POPOFF, MICHEL Y.

APPLICANT: LE GURN FELLOUS, MURIEL

TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE DETECTION OF

TITLE OF INVENTION: SALMONELLA

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/586,272

FILING DATE: 16-JAN-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9500410

FILING DATE: 16-JAN-1995

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 660-107-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

US-08-586-272-15

Query Match 64.0%; Score 12.8; DB 2; Length 20;

Best Local Similarity 87.5%; Pred No. 1.9e+02;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acggcagtgactgaat 17

Db 2 ACGGCATGGGCTGATT 17

RESULT 3

US-08-628-20/c

Sequence 20, Application US/09018628

Patent No. 5917019

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: van Steensel, Bas

APPLICANT: Bianchi, Alessandro

TITLE OF INVENTION: AN ALTERED TELOMERE REPEAT BINDING

TITLE OF INVENTION: FACTOR AND THERAPEUTIC USE THEREOF

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,378

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/018,628

FILING DATE:

ATTORNEY/AGENT INFORMATION:

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS

NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-191 CIP  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-09-273-378-20

Query Match 61.0%; Score 12.2; DB 5; Length 26;  
Best Local Similarity 82.4%; Pred. No. 4.1e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ggcagtgactgaatcg 20  
||||| ||||| ||  
Db 24 GGCATGCACTGAACAG 8

RESULT 5  
US-08-604-991-10/c  
; Sequence 10, Application US/08604991  
; Patent No. 6001565  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Andrew J.  
; APPLICANT: Jones, Dennis Mackay  
; TITLE OF INVENTION: Detection and Speciation of  
; TITLE OF INVENTION: Campylobacter  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
; STREET: 1100 New York Avenue, NW, Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604,991  
; FILING DATE: 29-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/GB94/01967  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9318751.6  
; FILING DATE: 09-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 1581.0040000  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-317-2540  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 41 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-604-991-10

Query Match 61.0%; Score 12.2; DB 5; Length 41;  
Best Local Similarity 82.4%; Pred. No. 4.4e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 aacggcagtgactgaat 17  
||| ||||| ||| ||||  
Db 36 AACTGCATTGACAGAAT 20

RESULT 6  
US-08-924-695A-28  
; Sequence 28, Application US/08924695A  
; Patent No. 5998583  
; GENERAL INFORMATION:  
; APPLICANT: KORSMEYER, STANLEY J.  
; TITLE OF INVENTION: BH3 INTERACTING DOMAIN DEATH AGONIST  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HOWELL & HAERKAMP, L.C.  
; STREET: 7733 FORSYTH BLVD., SUITE 1400  
; CITY: ST. LOUIS  
; STATE: MISSOURI  
; COUNTRY: USA  
; ZIP: 63105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/924,695A  
; FILING DATE: 09-SEP-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HOLLAND, DONALD R.  
; REGISTRATION NUMBER: 35,197  
; REFERENCE/DOCKET NUMBER: 971798  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314) 727-5188  
; TELEFAX: (314) 727-6092  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-924-695A-28

Query Match 59.0%; Score 11.8; DB 4; Length 36;  
Best Local Similarity 86.7%; Pred. No. 7e+02;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacggcagtgactga 15  
|| || ||||| ||||  
Db 22 AATGGGATGGACTGA 36

RESULT 7  
US-08-478-039-20  
; Sequence 20, Application US/08478039  
; Patent No. 5681722  
; GENERAL INFORMATION:  
; APPLICANT: Newman, Roland A.  
; APPLICANT: Hanna, Nabil  
; APPLICANT: Raab, Ronald W.  
; TITLE OF INVENTION: Recombinant Antibodies for Human Therapy  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

```
; STREET: 699 Prince St.
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,039
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/379,072
; FILING DATE: 25-JAN-1995
; APPLICATION NUMBER: US 07/912,292
; FILING DATE: 10-JUL-1992
; APPLICATION NUMBER: US 07/856,281
; FILING DATE: 23-MAR-1992
; APPLICATION NUMBER: US 07/735,064
; FILING DATE: 25-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin Esq., Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-160
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens or Monkey
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: heavy chain primers with Sali site
; US-08-478-039-20

Query Match 58.0%; Score 11.6; DB 1; Length 26;
Best Local Similarity 77.8%; Pred. No. 8.4e+02;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatggactgaatcgg 20
   ||| ||| ||| ||| |||
Db 8 CGACATGGACTGGACCTG 25

RESULT 8
US-08-476-349A-20
; Sequence 20, Application US/08476349A
; Patent No. 5750105
; GENERAL INFORMATION:
; APPLICANT: Newman, Roland A.
; APPLICANT: Hanna, Nabil
; APPLICANT: Raab, Ronald W.
; TITLE OF INVENTION: Recombinant Antibodies for Human Therapy
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince St.
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
```

```
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,349A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/379,072
; FILING DATE: 25-JAN-1995
; APPLICATION NUMBER: US 07/912,292
; FILING DATE: 10-JUL-1992
; APPLICATION NUMBER: US 07/856,281
; FILING DATE: 23-MAR-1992
; APPLICATION NUMBER: US 07/735,064
; FILING DATE: 25-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin Esq., Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens or Monkey
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: heavy chain primers with Sali site
; US-08-476-349A-20

Query Match 58.0%; Score 11.6; DB 2; Length 26;
Best Local Similarity 77.8%; Pred. No. 8.4e+02;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatggactgaatcgg 20
   ||| ||| ||| ||| |||
Db 8 CGACATGGACTGGACCTG 25

RESULT 9
US-08-859-998-657
; Sequence 657, Application US/08859998
; Patent No. 5994076
; GENERAL INFORMATION:
; APPLICANT: Chenchik, Alex
; APPLICANT: Jokhadze, George
; APPLICANT: Bibilashvili, Robert
; TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL
; NUMBER OF SEQUENCES: 1375
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: CA
; COUNTRY: US
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-451-240-42

```

```
Query Match          58.0%; Score 11.6; DB 1; Length 35;
Best Local Similarity 77.8%; Pred. No. 8.8e+02;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatgactgaatcgg 20
   | | | | | | | | | |
Db 26 CCGCATGATCGTATCGG 9

RESULT 12
US-08-470-846A-38/c
; Sequence 38, Application US/08470846A
; Patent No. 5846932
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Lowe, David G.
; APPLICANT: Cunningham, Brian C.
; APPLICANT: Oare, David
; APPLICANT: McDowell, Robert S.
; APPLICANT: Burnier, John
; TITLE OF INVENTION: RECEPTOR SPECIFIC ATRIAL NATRIURETIC PEPTIDES
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470, 846A
; FILING DATE: 06-Jun-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419877
; FILING DATE: 11-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362552
; FILING DATE: 06-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/152994
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kubinec, Jeffrey S.
; REGISTRATION NUMBER: 36,575
; REFERENCE/DOCKET NUMBER: P0844P2C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8228
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-470-846A-38

Query Match          58.0%; Score 11.6; DB 3; Length 35;
Best Local Similarity 77.8%; Pred. No. 8.8e+02;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatgactgaatcgg 20
   | | | | | | | | | |
Db 26 CCGCATGATCGTATCGG 9

Query Match          58.0%; Score 11.6; DB 6; Length 35;
Best Local Similarity 77.8%; Pred. No. 8.8e+02;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatgactgaatcgg 20
   | | | | | | | | | |
Db 26 CCGCATGATCGTATCGG 9
```

```
RESULT 13
PCT-US94-12591-42/c
; Sequence 42, Application PC/TUS9412591
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Lowe, David
; APPLICANT: Cunningham, Brian C.
; APPLICANT: Oare, David
; APPLICANT: McDowell, Robert S.
; APPLICANT: Burnier, John
; TITLE OF INVENTION: RECEPTOR SPECIFIC ATRIAL NATRIURETIC
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12591
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/152994
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Daryl B.
; REGISTRATION NUMBER: 32,637
; REFERENCE/DOCKET NUMBER: 844P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1249
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US94-12591-42

Query Match          58.0%; Score 11.6; DB 6; Length 35;
Best Local Similarity 77.8%; Pred. No. 8.8e+02;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 cggcatgactgaatcgg 20
   | | | | | | | | | |
Db 26 CCGCATGATCGTATCGG 9

RESULT 14
US-08-096-762-209/c
; Sequence 209, Application US/08096762
; Patent No. 5814318
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 5814318-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 200
```

CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,762  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/904,068  
FILING DATE: 23-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/853,408  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,279  
FILING DATE: 17-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 14643-9-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 209:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
US-08-096-762-209

Query Match 58.0%; Score 11.6; DB 2; Length 44;  
Best Local Similarity 77.8%; Pred. No. 9.1e+02;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cggcatggaactgaatcg 20  
||| ||||| |||  
Db 26 CGGGAAGGACTGACTCTG 9

RESULT 15  
US-08-041-538-13  
Sequence 13, Application US/08041538  
Patent No. 5385831  
GENERAL INFORMATION:  
APPLICANT: Mulvihill, Eileen R  
APPLICANT: Hagen, Frederick S  
APPLICANT: Houamed, Khaled M  
APPLICANT: Almers, Wolfhard  
TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE  
RECEPTORS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend  
STREET: Steuart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA

ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/041,538  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/648,481  
FILING DATE:  
APPLICATION NUMBER: US 07/626,806  
FILING DATE: 12-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 13952-6-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-467-9600  
TELEFAX: 206-623-6793  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
IMMEDIATE SOURCE:  
CLONE: ZC3016  
US-08-041-538-13

Query Match 57.0%; Score 11.4; DB 1; Length 21;  
Best Local Similarity 92.3%; Pred. No. 1e+03;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 acggcatggactg 14  
| ||||| |||||  
Db 5 ATGGCATGGACTG 17

Search completed: May 23, 2000, 11:19:51  
Job time: 5948 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 07:13:44 ; Search time 693.2 Seconds  
(without alignments)  
-28.067 Million cell updates/sec

Title: US-08-945-805-1  
Perfect score: 20  
Sequence: 1 ccttgaaggattccctcc 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_bal.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_bal.\*
- 37: em\_bal.\*
- 38: em\_hum3.\*
- 39: gb\_pr4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	20	100.0	992	12	MMVCAM1B1	U12878 Mus musculu
2	20	100.0	2321	12	MUSVCAM01	L22301 Mus musculu
3	20	100.0	2355	10	S50587	S50587 VCAM1-vascu
4	20	100.0	2396	9	HUMVCAMA	M92431 Human vascu
5	20	100.0	2458	12	MMU42327	U42327 Mus musculu
6	20	100.0	5607	9	HUMVCAM1A	M73255 Human vascu
7	17.4	87.0	148608	32	AP001124	AP001124 Homo sapi
8	17	85.0	136000	33	HSF10178	AL035367 Homo sapi
9	17	85.0	144596	44	AC020567	AC020567 Homo sapi
c 10	16.8	84.0	145	12	MMCREBG	X67724 Mus musculu
c 11	16.8	84.0	436	13	G25297	G25297 human STS E
c 12	16.8	84.0	436	13	G29465	G29465 human STS S
c 13	16.8	84.0	625	8	CNS019F5	AL11561 Botrytis
c 14	16.8	84.0	739	11	HSJ243671	AJ243671 Homo sapi
c 15	16.8	84.0	1007	12	MMU46027	U46027 Mus musculu
c 16	16.8	84.0	1023	12	RNCREB	X60002 Rat delta C
c 17	16.8	84.0	1125	12	RNCREB	X14788 Rat mRNA fo
c 18	16.8	84.0	1161	12	MMTRANS	X92497 M.musculus
c 19	16.8	84.0	1258	12	MUSCREB	M95106 Mus musculu
c 20	16.8	84.0	5414	14	AF049616	AF049616 Cloning v
c 21	16.8	84.0	5414	14	AF063583	AF063583 Cloning v
c 22	16.8	84.0	77617	43	AC016521	AC016521 Homo sapi
c 23	16.8	84.0	87705	11	HSJ36315	AL049821 Human DNA
c 24	16.8	84.0	106902	56	AC011430	AC011430 Homo sapi
c 25	16.8	84.0	151469	45	AC018514	AC018514 Homo sapi
c 26	16.8	84.0	170680	45	AC011201	AC011201 Homo sapi
c 27	16.8	84.0	176373	44	AC018373	AC018373 Homo sapi
c 28	16.8	84.0	186379	52	AC022182	AC022182 Homo sapi
c 29	16.8	84.0	197199	33	AC005054	AC005054 Homo sapi
c 30	16.8	84.0	233932	54	AC009175	AC009175 Homo sapi
c 31	16.8	84.0	244595	41	AC008117	AC008117 Homo sapi
c 32	16.4	82.0	4442	1	ATAGROSY	Z29717 A.tumefacie
c 33	16.4	82.0	146174	1	D90910	D90910 Synecocyst
c 34	16.4	82.0	151670	41	AC007777	AC007777 Homo sapi
c 35	16.4	82.0	151906	45	AC019326	AC019326 Mus muscu
c 36	16.4	82.0	169511	41	AC007717	AC007717 Homo sapi
c 37	16.4	82.0	174571	33	AL138764	AL138764 Homo sapi
c 38	16.4	82.0	180943	11	CNS01DV1	AL133445 Human chr
c 39	16.4	82.0	191083	32	HSJ776F14	AL132670 Homo sapi
c 40	16.4	82.0	255548	33	HSJ776F14	AL109658 Homo sapi
c 41	16	80.0	1175	2	LGU91638	U91638 Legionella
c 42	16	80.0	33539	35	CELF54E7	U00067 Caenorhabdi
c 43	16	80.0	41594	41	AC010964	AC010964 Homo sapi
c 44	16	80.0	122228	11	AC000003	AC000003 Homo sapi
c 45	16	80.0	150208	33	AL139391	AL139391 Homo sapi

ALIGNMENTS

```

RESULT 1
LOCUS MVVCAM1B1 992 bp 28-OCT-1995
DEFINITION Mus musculus NIH Swiss vascular cell adhesion molecule-1 (VCAM-1)
ACCESSION U12878
VERSION U12878.1 GI:1041794
KEYWORDS
SOURCE 1 of 6
SEGMENT mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 992)
AUTHORS Kumar, A.G., Dai, Y.X., Kozak, C.A., Mims, M.P., Gotto, A.M. and
Ballantyne, C.M.
TITLE Murine VCAM-1: Molecular cloning, Mapping, and Analysis of a
Truncated Form
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 992)
AUTHORS Kumar, A.G.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-1994) Ajith G. Kumar, Department of Medicine,
Baylor College of Medicine, 6535 Fannin Street, Houston, TX 77030,
USA
COMMENT On Oct 28, 1995 this sequence version replaced gi:531856.
FEATURES
source
Location/Qualifiers
1..992
/organism="Mus musculus"
/strain="NIH Swiss"
/db_xref="taxon:10090"
/clone_lib="lambda 1"
/clone_lib="Genomic library lambda FIX II, Stratagene, La
Jolla, CA"
/chromosome="3"
/cell_line="NIH3T3"
/dev_stage="fibroblast"
TATA_signal
sig_peptide 632..636
exon 751..814
CDS <751..814
/gene="VCAM-1"
/number=1
751..>814
/gene="VCAM-1"
/codon_start=1
/product="vascular cell adhesion molecule-1"
/protein_id="AAB60659.1"
/db_xref="GI:531863"
/translation="MPVKRVAVLGASTVWILFAV"
BASE COUNT 280 a 204 c 204 g 304 t
ORIGIN
Query Match 100.0%; Score 20; DB 12; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctcc 20
|||||
Db 592 CCTTGAAGGATTCCCTCC 611

RESULT 2
LOCUS MUSVCAM01 2321 bp 23-MAR-1994
DEFINITION Mus musculus vascular cell adhesion molecule-1 (VCAM1) gene, exons
1 and 2.
ACCESSION L22301
VERSION L22301.1 GI:347974
KEYWORDS vascular cell adhesion molecule-1.

```

```

SEGMENT 1 of 8
SOURCE Mus musculus strain 129 DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2321)
AUTHORS Cybulsky, M.I., Allan-Motamed, M. and Collins, T.
TITLE Structure of the murine VCAM1 gene
JOURNAL Genomics 18 (2), 387-391 (1993)
MEDLINE 94117008
COMMENT MVI-6 individual isolate with similarities to MVI-9 and MVI-3; exon
1 encodes the signal peptide, and exon 2 encodes the first
immunoglobulin-like domain of murine VCAM-1.
FEATURES
source
Location/Qualifiers
1..2321
/organism="Mus musculus"
/isolate="MVI-7"
/strain="129"
/sub_species="domesticus"
/isolate="MVI-9"
/isolate="MVI-6"
/db_xref="taxon:10090"
/cell_line="J1"
/cell_type="stem cells"
/dev_stage="blastula (blastocyst)"
/germline
/tissue_type="embryo"
/tissue_lib="lambda DASH II"
1..1340
/promoter
TATA_signal
5'UTR 1341..1434
exon /gene="VCAM1"
1341..1498
/intron /number=1
1499..1978
exon /gene="VCAM1"
1979..2254
/number=2
2255..2321
/number=2
BASE COUNT 688 a 467 c 469 g 697 t
ORIGIN
Query Match 100.0%; Score 20; DB 12; Length 2321;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctcc 20
|||||
Db 1277 CCTTGAAGGATTCCCTCC 1296

RESULT 3
LOCUS S50587 2355 bp DNA
DEFINITION VCAM1=vascular cell adhesion molecule 1 [5' region, promoter]
[human, Genomic, 2355 nt].
ACCESSION S50587
VERSION S50587.1 GI:261284
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2355)
AUTHORS Neish, A.S., Williams, A.J., Palmer, H.J., Whitley, M.Z. and Collins, T.
TITLE Functional analysis of the human vascular cell adhesion molecule 1
promoter
JOURNAL J. Exp. Med. 176 (6), 1583-1593 (1992)

```

```

MEDLINE 93094762
REMARK  GenBank staff at the National Library of Medicine created this
        entry [NCBI gibsseq 120211] from the original journal article.
        This sequence comes from Fig. 1.
        Map location: lp31-32.
FEATURES             Location/Qualifiers
     source           1. .2355
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene             1. .2355
                     /partial
                     /gene="VCAM1"
                     /note="vascular cell adhesion molecule 1"
BASE COUNT          728 a 447 c 431 g 746 t
ORIGIN
1 ccttgaaggatttcctcc 20
|||||
Db 2104 CCTTGAAGGATTTCCCTCC 2123

Query Match          100.0%; Score 20; DB 10; Length 2355;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcctcc 20
|||||
Db 2104 CCTTGAAGGATTTCCCTCC 2123

RESULT 4
LOCUS HUMVCAM1 2396 bp DNA PRI 14-JAN-1995
DEFINITION Human vascular cell adhesion molecule-1 (VCAM1) gene, exon 1.
ACCESSION M92431
VERSION M92431.1 GI:340197
KEYWORDS vascular cell adhesion molecule-1.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 2396)
JOURNAL Iademarco,M.F., McQuillan,J.J., Rosen,G.D. and Dean,D.C.
MEDLINE Characterization of the promoter for vascular cell adhesion
          molecule-1 (VCAM-1)
          J. Biol. Chem. 267 (23), 16323-16329 (1992)
FEATURES             Location/Qualifiers
     source           1. .2396
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     promoter          1. .2180
                     /map="lp32-p31"
                     /gene="VCAM1"
                     /note="G00-127-922"
     gene              1. .2396
                     /gene="VCAM1"
                     /note="G00-127-922"
     TATA_signal        2152. .2157
                     /gene="VCAM1"
     exon              2300. .2363
                     /note="G00-127-922"
                     /gene="VCAM1"
     intron             2364. .>2396
                     /number=1
                     /note="G00-127-922"
                     /number=1
BASE COUNT           740 a 454 c 436 g 766 t
ORIGIN
1 ccttgaaggatttcctcc 20

Query Match          100.0%; Score 20; DB 9; Length 2396;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcctcc 20

```

```

|||||
Db 2112 CCTTGAAGGATTTCCCTCC 2131

RESULT 5
LOCUS MMU42327 2458 bp DNA ROD 09-DEC-1997
DEFINITION Mus musculus vascular adhesion molecule-1 (VCAM-1) gene, partial
          cds.
ACCESSION U42327
VERSION U42327.1 GI:1147764
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          1 (bases 1 to 2458)
          Korenaga,R., Ando,J., Kosaki,K., Isshiki,M., Takada,Y. and
          Kamiya,A.
TITLE Negative transcriptional regulation of the VCAM-1 gene by fluid
          shear stress in murine endothelial cells
JOURNAL Am. J. Physiol. 273 (5), C1506-C1515 (1997)
MEDLINE 98042048
REFERENCE 2 (bases 1 to 2458)
AUTHORS Korenaga,R., Ando,J., Tsuboi,H. and Kamiya,A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-1995) Risa Korenaga, Cardiovascular Biomechanics,
          Faculty of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku,
          Tokyo 113, Japan
FEATURES             Location/Qualifiers
     source           1. .2458
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_type="vascular endothelial cell"
                     1896. .1958
                     /gene="VCAM-1"
                     1896. .>1958
                     /gene="VCAM-1"
                     /codon_start=1
                     /product="vascular adhesion molecule-1"
                     /protein_id="AAB88576.1"
                     /db_xref="GI:1147765"
                     /translation="MPVKMVAVLGASTVLWILFAV"
BASE COUNT           718 a 482 c 482 g 776 t
ORIGIN
1 ccttgaaggatttcctcc 20
|||||
Db 1738 CCTTGAAGGATTTCCCTCC 1757

Query Match          100.0%; Score 20; DB 12; Length 2458;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcctcc 20
|||||
Db 1738 CCTTGAAGGATTTCCCTCC 1757

RESULT 6
LOCUS HUMVCAM1A 5607 bp DNA PRI 14-JAN-1995
DEFINITION Human vascular cell adhesion molecule-1 (VCAM1) gene, complete CDS.
ACCESSION M73255
VERSION M73255.1 GI:340195
KEYWORDS Immunoglobulin super gene family; leukocyte adhesion molecule;
          transmembrane protein; vascular cell adhesion molecule-1.
          Homo sapiens (tissue library: EMBL3) peripheral blood DNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 5607)
          Cybulsky,M.I., Fries,J.W., Williams,A.J., Sultan,P., Eddy,R.,
          Byers,M., Shows,T., Gimbrone,M.A. Jr. and Collins,T.
TITLE Gene structure, chromosomal location, and basis for alternative
          mRNA splicing of the human VCAM1 gene

```



34904 52029 contig of 17126 bp in length  
52530 68671 contig of 16142 bp in length  
69172 82125 contig of 12954 bp in length  
82626 95169 contig of 12544 bp in length  
95670 102923 contig of 7254 bp in length  
103424 112232 contig of 8809 bp in length  
112733 119833 contig of 7101 bp in length  
120334 128452 contig of 8119 bp in length  
128953 134016 contig of 5064 bp in length  
134517 139515 contig of 4999 bp in length  
140016 143347 contig of 3332 bp in length  
143848 146313 contig of 2466 bp in length  
146814 148608 contig of 1795 bp in length

\* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 16639: contig of 16639 bp in length  
16640 17139: gap of 500 bp  
17140 34402: contig of 17263 bp in length  
34403 34905: gap of 503 bp  
34906 52028: contig of 17123 bp in length  
52029 52530: gap of 502 bp  
52531 68671: contig of 16141 bp in length  
68672 69176: gap of 505 bp  
69177 82125: contig of 12949 bp in length  
82126 82627: gap of 502 bp  
82628 95168: contig of 12541 bp in length  
95169 95669: gap of 501 bp  
95670 102922: contig of 7253 bp in length  
102923 103430: gap of 508 bp  
103431 112231: contig of 8801 bp in length  
112232 112733: gap of 502 bp  
112734 119832: contig of 7099 bp in length  
119833 120335: gap of 503 bp  
120336 128452: contig of 8117 bp in length  
128453 128952: gap of 500 bp  
128953 134016: contig of 5064 bp in length  
134017 134516: gap of 500 bp  
134517 139514: contig of 4998 bp in length  
139515 140015: gap of 501 bp  
140016 143346: contig of 3331 bp in length  
143347 143847: gap of 501 bp  
143848 146310: contig of 2463 bp in length  
146311 146815: gap of 505 bp  
146816 148608: contig of 1793 bp in length.

## FEATURES

Source  
1. 148608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="RP11-775A1"  
/map="11q23"

BASE COUNT 43553 a 27859 c 28047 g 41248 t 7901 others  
ORIGIN

Query Match 87.0%; Score 17.4; DB 32; Length 148608;  
Best Local Similarity 94.7%; Pred. No. 54;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccttgaaggaggatttcctc 19  
|||||

Db 83083 CTTGAAGGGATTCTCTC 83101

## RESULT 8

HSF10178 136000 bp DNA HTG 27-JAN-2000  
LOCUS

## DEFINITION

Homo sapiens chromosome 17 clone PAC RPC1-1 178F10 map 17p11.2, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 5 unordered pieces.

## ACCESSION

AL035367

## VERSION

AL035367.1 GI:6982189

## KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

## REFERENCE

1 (bases 1 to 136000)  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

## AUTHORS

Radelof, U., Hennig, S., Ramser, J., Francis, F., Steffens, C.,  
Klein, M., Seranski, P., Poustka, A., Reinhard, R. and Lehrach, H.

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 136000)

## AUTHORS

MPIMG.

## TITLE

Direct Submission

## JOURNAL

Submitted (05-FEB-1999) MPIMG, Abt. Lehrach, Max Planck Institut  
Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany  
On Feb 16, 2000 this sequence version replaced gi:6729615.

## COMMENT

contig 01 1. 16754  
contig 02 17555..26092  
contig 03 26893..72067  
contig 04 72868..123726  
contig 05 124527..136000

Clone received from the Resource Centre of the Human Genome Project  
at the Max-planck-Institute for Molecular Genetics.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 16754: contig of 16754 bp in length  
\* 17555 17554: gap of 800 bp  
\* 17555 26092: contig of 8538 bp in length  
\* 26093 26892: gap of 800 bp  
\* 26893 72067: contig of 45175 bp in length  
\* 72068 72867: gap of 800 bp  
\* 72868 123726: contig of 50859 bp in length  
\* 123727 124526: gap of 800 bp  
\* 124527 136000: contig of 11474 bp in length.

Location/Qualifiers  
1. 136000  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17p11.2"

/clone="PAC RPC1-1 178F10"  
/clone\_lib="RPC11,3-5 Human PAC library, originating  
insitute: Roswell Park Cancer Institute, creator: Pletzer  
de Jong, P. Ioannou"

/note="region between markers D17S842-D17S953"  
BASE COUNT 31167 a 35065 c 35050 g 31518 t 3200 others  
ORIGIN

Query Match 85.0%; Score 17; DB 33; Length 136000;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttgaaggaggatttcctc 17  
|||||

Db 14900 CTTGAAGGGATTTCCTC 14916

RESULT 9  
AC020567 144596 bp DNA HTG 04-JAN-2000  
LOCUS

AC020567 Homo sapiens chromosome UL clone RP11-189D22, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 47 unordered pieces.

AC020567

VERSION	AC020567.1	GI:6670874	
KEYWORDS	HTG: HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 144596)		
TITLE	Waterston,R.H.		
JOURNAL	The sequence of Homo sapiens clone		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 144596)		
JOURNAL	Waterston,R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (04-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	Center project name: H_NH0189D22. * NOTE: This is a 'working draft' sequence. It currently consists of 47 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.		
	1	2164:	contig of 2164 bp in length
		2165:	gap of unknown length
		2184:	contig of 1960 bp in length
		4144:	gap of unknown length
		4162:	contig of 2235 bp in length
		6397:	contig of 2235 bp in length
		6398:	gap of unknown length
		6416:	contig of 1947 bp in length
		8362:	gap of unknown length
		8364:	contig of 2606 bp in length
		8383:	contig of 1972 bp in length
		10989:	contig of 1972 bp in length
		11007:	contig of 1972 bp in length
		12978:	contig of 1972 bp in length
		12980:	contig of 1972 bp in length
		12989:	contig of 1972 bp in length
		15563:	contig of 2566 bp in length
		15565:	gap of unknown length
		17934:	contig of 2351 bp in length
		17935:	gap of unknown length
		17953:	contig of 2002 bp in length
		17954:	contig of 2002 bp in length
		19956:	contig of 2002 bp in length
		19974:	contig of 2002 bp in length
		22182:	contig of 2208 bp in length
		22201:	contig of 2208 bp in length
		22538:	contig of 3337 bp in length
		25539:	contig of 3337 bp in length
		25557:	contig of 3337 bp in length
		25558:	contig of 2521 bp in length
		28079:	contig of 2521 bp in length
		30400:	contig of 2303 bp in length
		30401:	contig of 2303 bp in length
		30420:	contig of 2600 bp in length
		33019:	contig of 2600 bp in length
		33038:	contig of 1999 bp in length
		33039:	contig of 1999 bp in length
		35037:	contig of 1999 bp in length
		35038:	contig of 1999 bp in length
		35056:	contig of 1999 bp in length
		38051:	contig of 2995 bp in length
		38052:	contig of 2995 bp in length
		38070:	contig of 3384 bp in length
		41454:	contig of 3384 bp in length
		41473:	contig of 3384 bp in length
		44312:	contig of 2839 bp in length
		44313:	contig of 2839 bp in length
		44332:	contig of 2211 bp in length
		46542:	contig of 2211 bp in length
		46561:	contig of 2211 bp in length
		46562:	contig of 2175 bp in length
		48737:	contig of 2175 bp in length
		48755:	contig of 2175 bp in length
		51140:	contig of 2385 bp in length
		51141:	contig of 2385 bp in length
		51159:	contig of 2385 bp in length
		53517:	contig of 2358 bp in length
		53518:	contig of 2358 bp in length
		53537:	contig of 2391 bp in length
		55928:	contig of 2391 bp in length
		55947:	contig of 3148 bp in length
		59095:	contig of 3148 bp in length
		59113:	contig of 3148 bp in length
		59114:	contig of 2936 bp in length
		62050:	gap of unknown length
		62056:	contig of 2988 bp in length
		65075:	gap of unknown length
		68489:	contig of 3414 bp in length
		68490:	gap of unknown length
		68509:	contig of 4666 bp in length
		73174:	contig of 4666 bp in length
		73193:	contig of 4666 bp in length
		76800:	contig of 3607 bp in length
		76819:	contig of 3607 bp in length
		79917:	contig of 3098 bp in length
		79936:	contig of 3098 bp in length
		83029:	contig of 3093 bp in length
		83048:	contig of 3093 bp in length
		85338:	contig of 2490 bp in length
		85357:	contig of 2490 bp in length
		88642:	contig of 3085 bp in length
		88661:	contig of 3085 bp in length
		91603:	contig of 2942 bp in length
		91622:	contig of 2942 bp in length
		94830:	contig of 3208 bp in length
		94849:	contig of 3208 bp in length
		97795:	contig of 2946 bp in length
		97814:	contig of 2946 bp in length
		101185:	contig of 3371 bp in length
		101204:	contig of 3371 bp in length
		106074:	contig of 4870 bp in length
		106093:	contig of 4870 bp in length
		108910:	contig of 2798 bp in length
		108991:	contig of 2798 bp in length
		108992:	contig of 2798 bp in length
		108911:	contig of 3876 bp in length
		112787:	contig of 3876 bp in length
		112805:	contig of 3876 bp in length
		117388:	contig of 4583 bp in length
		117389:	contig of 4583 bp in length
		123007:	contig of 5600 bp in length
		123008:	contig of 5600 bp in length
		123027:	contig of 5292 bp in length
		128319:	contig of 5292 bp in length
		128338:	contig of 5292 bp in length
		134777:	contig of 5292 bp in length
		134796:	contig of 5292 bp in length
		140055:	contig of 5260 bp in length
		140073:	contig of 5260 bp in length
		140074:	contig of 5260 bp in length
		142095:	contig of 2022 bp in length
		142096:	contig of 2022 bp in length
		142114:	contig of 2483 bp in length.
FEATURES	Location/Qualifiers		
source	1..144596		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="UL"		
	/clone="RP11-189D22"		
BASE COUNT	36240 a	35780 c	34034 g 37670 t 872 others
ORIGIN			
Query Match	85.0%	Score 17:	DB 44; Length 144596;
Best Local Similarity	100.0%	Pred. NO. 90;	
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	ccttgaaggattcccc	17
Db	74604	CCTTGAAGGATTCCC	74620
RESULT	10		
MMCREBG/c			
LOCUS	MMCREBG	145 bp	DNA
DEFINITION	Mus musculus CREB gene for CAMP-responsive-element binding protein, exon 7.		14-JUL-1997
ACCESSION	X67724	X65763	
VERSION	X67724.1	GI:288940	
KEYWORDS	CAMP responsive element binding protein; creB gene; sequence-specific binding protein.		

SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 145)  
 AUTHORS Ruppert, S.J.W.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-AUG-1992) S.J.W. Ruppert, Howard Hughes Medical  
 Institute, University of California at Berkeley, Dept of Molecular  
 & Cell Biology, 401 Barker Hall, Berkeley, CA 94720, USA  
 REFERENCE 2 (bases 1 to 145)  
 AUTHORS Ruppert, S., Cole, T.J., Boshart, M., Schmid, E. and Schutz, G.  
 TITLE Multiple mRNA isoforms of the transcription activator protein CREB:  
 generation by alternative splicing and specific expression in  
 primary spermatocytes  
 JOURNAL EMBO J. 11 (4), 1503-1512 (1992)  
 MEDLINE 92224889  
 COMMENT Related sequences M95106, M95107 and X67718-X67728.  
 FEATURES  
 source  
 1. .145  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="c-ch/c-ch"  
 /db\_xref="taxon:10090"  
 /cell\_line="melanoma B16, teratocarcinoma F9"  
 /map="2q32.3-q34"  
 1. .145  
 /gene="CREB"  
 <1. .28  
 /gene="CREB"  
 /number=6  
 /evidence=experimental  
 29. .129  
 /gene="CREB"  
 /number=7  
 /usedin=X67719:crebd\_cds  
 /usedin=X67719:creba\_cds  
 /usedin=X67719:crebw\_cds  
 /usedin=X67718:crebd\_mRNA  
 /usedin=X67718:creba\_mRNA  
 /usedin=X67718:crebag\_mRNA  
 /usedin=X67718:crebgsi\_mRNA  
 /usedin=X67718:crebpsi\_mRNA  
 /evidence=experimental  
 130. .>145  
 /gene="CREB"  
 /number=7  
 /evidence=experimental  
 47 a 23 c 31 g 44 t  
 BASE COUNT  
 ORIGIN  
 Query Match 84.0%; Score 16.8; DB 12; Length 145;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 ccttgaaggatttcctccc 20  
 ||||| ||||| |||||  
 Db 115 CCTTGAAGGATTTCCCTTC 96  
 RESULT 11  
 G25297/c 436 bp DNA STS 02-JUN-1996  
 LOCUS human STS EST163198, sequence tagged site.  
 DEFINITION G25297  
 ACCESSION G25297  
 VERSION G25297.1 GI:1347529  
 KEYWORDS STS; STS sequence; primer; sequence tagged site.  
 SOURCE human STS derived from sequences in dbEST and the Unigene  
 collection.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 436)  
 AUTHORS  
 JOURNAL  
 COMMENT

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Hudson, T.  
 Whitehead Institute/MIT Center for Genome Research; Physically  
 Mapped STSs  
 Unpublished (1995)  
 Contact: Thomas Hudson  
 Whitehead Institute/MIT Center for Genome Research  
 Whitehead Institute for Biomedical Research  
 9 Cambridge Center, Cambridge MA 02142 USA  
 Tel: 617 252 1900  
 Fax: 617 252 1902  
 Email: thudson@genome.wi.mit.edu

Primer A: ATCTCTCTCATTTTCCACCC  
 Primer B: TCACAACCTGGTGACTGGA  
 STS size: 127  
 PCR Profile:  
 Presoak:  
 Denaturation:  
 Annealing: 56 degrees C  
 Polymerization:  
 PCR Cycles: 35  
 Thermal Cycler:

Protocol:  
 Template: 10 ng  
 Primer: each 5 pM  
 dNTPs: each 4 nM  
 Taq Polymerase: 0.025 units/ul  
 Total Vol: 20 ul

Buffer:  
 MgCl2: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 9.3

Derived from dbEST (genbank accession T93077).

FEATURES  
 source

1. .436  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="726.9 CR from top of Chr1 linkage group"

STS  
 21. .147

primer\_bind 21. .40

primer\_bind complement(128..147)

BASE COUNT 121 a 90 c 98 g 124 t 3 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 436;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ccttgaaggatttcctccc 20

||||| ||||| |||||

Db 184 CCTTGAGGGATTTCCCTCC 165

RESULT 12

G29465/c

LOCUS human STS SHGC-33740, sequence tagged site.

DEFINITION G29465

ACCESSION G29465

VERSION G29465.1 GI:1593016

KEYWORDS STS; STS sequence; primer; sequence tagged site.

SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 436)

AUTHORS

JOURNAL

COMMENT Unpublished (1996)

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu

Primer A: ATCCCTCTCCATTTTCCACCC  
Primer B: TCACAACTGCTGACTGAA  
STS size: 127  
PCR Profile:

Initial incubation: 94 degrees C for 90 seconds  
Denaturation: 94 degrees C for 15 seconds  
Annealing: 62 degrees C for 23 seconds  
Polymerization: 72 degrees C for 30 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9600  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Taq Polymerase: 0.05 units/ul  
Total Vol: 10 ul  
Buffer: MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 20 mM  
pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from T93077  
-- Washington University/Merck EST sequence.

FEATURES  
source  
1. .436  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="1"  
21. .147  
21. .40

STS  
primer\_bind 21. .40  
primer\_bind complement(128. .147)  
BASE COUNT 121 a 90 c 98 g 124 t 3 others  
ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 436;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcctcc 20  
||||| ||||||| |||  
Db 184 CCTGAGGGGATTCCTTC 165

RESULT 13  
CNS019F5 625 bp mRNA PLN 02-SEP-1999  
LOCUS Botrytis cinerea strain T4 cdna library under conditions of  
DEFINITION nitrogen deprivation.  
ACCESSION AL11561  
VERSION AL11561.1 GI:5826180  
KEYWORDS cdna library; nitrogen deprivation.  
SOURCE Botryotinia fuckeliana.  
ORGANISM Botryotinia fuckeliana  
Eukaryota; Fungi; Ascomycota; Euascomycetes; Discomycetes;  
Leotiales; Sclerotiniaceae; Botryotinia.  
1 (bases 1 to 625)  
REFERENCE  
AUTHORS Bitton.F., Lewis.C., Fortini.D., Pradler.J.M. and Brygoo.Y.  
TITLE Direct Submission  
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,  
78026 Versailles, France  
REFERENCE 2 (bases 1 to 625)

AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :  
CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT The cdna library to be analyzed within the framework of this  
project was created using a Botrytis cinerea strain which was grown  
under conditions of nitrogen deprivation, which is the normal  
situation for B. cinerea during its development on its host plant.  
The library was produced in an oriented direction, in the pBSII  
vector.

FEATURES  
source  
1. .625  
Location/Qualifiers  
/organism="Botryotinia fuckeliana"  
/strain="T4"  
/db\_xref="taxon:40559"  
/note="Genoscope sequence ID : W20H101"  
BASE COUNT 171 a 138 c 148 g 168 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 8; Length 625;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcctcc 20  
||||| ||||||| |||||||  
Db 26 CTTTCAAGGGATTCCTCC 45

RESULT 14  
HSA243671 739 bp mRNA PRI 07-JAN-2000  
LOCUS Homo sapiens partial mRNA for NICE-4 protein, 3' end, clone  
DEFINITION 3114f17.  
ACCESSION AJ243671  
VERSION AJ243671.1 GI:6688160  
KEYWORDS NICE-4 gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 739)  
AUTHORS Marenholz,I., Zirra,M., Fischer,D.F., Backendorf,C., Ziegler,A. and  
Mischke,D.  
TITLE Novel genes in the Human Epidermal Differentiation Complex (EDC)  
identified by subtractive hybridization of entire YACs to a gridded  
keratinocyte cDNA library  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 739)  
AUTHORS Marenholz,I.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUN-1999) Marenholz I., Universitaetsklinikum  
Charite, Humboldt-Universitaet Berlin, Institut fuer Immunogenetik,  
Spandauer Damm 130, 14050 Berlin, GERMANY  
COMMENT Related sequences: D63478,AJ243668,AJ243669,AJ243670 (splice  
variants)  
Related sequences: AA01857,AA361152 (overlapping clones).

FEATURES  
source  
1. .739  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone\_lib="KER 3/4"  
/clone="3114f17"  
/map="1q21"  
1. .144  
/gene="NICE-4"  
<1. .144  
/gene="NICE-4"  
/codon\_start=1  
/product="NICE-4 protein"  
/protein\_id="CAB65102.1"

/db\_xref="GI:6688161"  
/translation="ATAAAYPPAPFPMHILTPHQPHSQILHHHLQDGDILNFVDDQ  
LGE"  
735  
polyA\_site 232 a 179 c 162 g 166 t  
BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 739;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ccttgaaggattccctcc 20  
||||| ||||||| |||  
Db 256 CCTGAGGGGATTCTCTCC 275

RESULT 15  
MMU46027/c  
LOCUS  
DEFINITION Mus musculus CREB transcription factor, novel spliced form, mRNA, 25-JUL-1997  
partial cds.  
ACCESSION U46027  
VERSION U46027.1 GI:1655804  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1007)  
AUTHORS Yang,L., Lanier,E.R. and Kraig,E.  
TITLE Identification of a novel, spliced variant of CREB that is preferentially expressed in the thymus  
J. Immunol. 158 (6), 2522-2525 (1997)  
JOURNAL 97211815  
MEDLINE 2 (bases 1 to 1007)  
AUTHORS Kraig,E.  
TITLE Direct Submission  
JOURNAL Submitted (12-JAN-1996) Ellen Kraig, Univ. of Texas Health Science Center at San Antonio, Cellular & Structural Biology, 7703 Floyd Curl Dr., San Antonio, TX 78284, USA  
LOCATION/Qualifiers  
FEATURES  
source  
1..1007  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="pmly2"  
/tissue\_type="thymus"  
<1..774  
/note="a novel CREB family member; CAMP response element-binding protein"  
/codon\_start=1  
/function="transcription factor"  
/product="CREB"  
/protein\_id="AAB64015.1"  
/db\_xref="GI:1655805"  
/translation="EEFGTSAVTEAENQMTVQAQPIATLAQVMPAAHATSSAPTV  
TLVQLPNGRQNGHGVIOAAQPSVIOQSVQTVQISTAEESQESVDSVDSOKRR  
EILSRFPYRKILNDLSDAPGVARIEEKESEETSSPAITVTVPPIYQTSQGYT  
ASGDVQYQITAPTSTIAPGVMASSPALPTQPAEEAARKREVLKMKREARECRR  
KKKEYVKLENRVAVLEQNKTLEELKALKDLYCHKSD"

CDS  
316 a 244 c 215 g 232 t  
BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 1007;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ccttgaaggattccctcc 20  
||||| ||||||| |||  
Db 321 CCTGAAGGATTCTCCTC 302

Search completed: May 23, 2000, 09:39:11  
Job time: 8727 sec



Result No.	Score	Query Match	§			Description
			Length	DB	ID	
1	20	100.0	20	1	T59675	
2	20	100.0	36	1	Q90101	
3	20	100.0	47	1	Q90100	
4	20	100.0	1032	1	Q06690	
5	18	90.0	19	1	Q90103	
6	15.8	79.0	1548	1	N80038	
7	15.8	79.0	1549	1	N80398	
8	15.4	77.0	387	1	T67638	
9	15.4	77.0	387	1	T68240	
10	15.4	77.0	477	1	R86164	
11	15.4	77.0	1629	1	T18011	
12	15.2	76.0	316	1	T22912	
13	15.2	76.0	6171	1	V52170	
14	15	75.0	34	1	Q90105	
15	15	75.0	1230	1	T39170	
16	15	75.0	1860	1	Q26728	
17	14.8	74.0	534	1	T19067	
18	14.8	74.0	1438	1	T79126	
19	14.8	74.0	1526	1	V07152	
20	14.8	74.0	2475	1	Q04780	
21	14.8	74.0	3214	1	X13217	
22	14.4	72.0	485	1	X51803	
23	14.4	72.0	2313	1	V09330	
24	14.4	72.0	2313	1	V19347	
25	14.4	72.0	2352	1	Q74275	
26	14.4	72.0	2352	1	V09333	
27	14.4	72.0	2352	1	X04435	
28	14.4	72.0	2427	1	T070435	
29	14.4	72.0	2523	1	T62101	
30	14.4	72.0	2523	1	V01457	
31	14.4	72.0	2651	1	Q74268	
32	14.4	72.0	2651	1	X04428	
33	14.4	72.0	8357	1	T91532	
34	14.4	72.0	10993	1	V52148	

```

PI Bennett CF, Medford RM;
DR WPI: 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17; Page 33; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 36 BP; 3 A; 13 C; 9 G; 11 T;

Query Match 100.0%; Score 20; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggaggattccctcc 20
Db 16 CCTGAAGGGATTCCCTCC 35

RESULT 3
Q90100
ID Q90100 standard; DNA; 47 BP.
AC Q90100;
DT 11-JAN-1996 (first entry)
DE VCAM-1 expression inhibiting oligonucleotide.
KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;
KW transcriptional regulatory factor; diagnosis; treatment; restenosis;
KW atherosclerosis; inflammatory disease; ds.
OS Synthetic.
PN WO9512415-A1.
PD 11-MAY-1995.
PF 07-NOV-1994; U12797.
PR 05-NOV-1993; US-147878.
PA (ISTS-) ISIS PHARM INC.
PA (OYEM-) UNIV EMORY.
PI Bennett CF, Medford RM;
DR WPI: 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17; Page 33; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 47 BP; 3 A; 17 C; 12 G; 15 T;

Query Match 100.0%; Score 20; DB 1; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggaggattccctcc 20
Db 22 CCTGAAGGGATTCCCTCC 41

RESULT 4
Q06690
ID Q06690 standard; DNA; 1032 BP.
AC Q06690;
DT 27-FEB-1991 (first entry)
DE 5' UTR of VCAM1 from clone vcl1-16.
KW Endothelial cell-leucocyte adhesion molecule 1; ELAM1;
KW vascular cell adhesion molecule 1; VCAM1; UTR; vector; promoter;
KW molecule involved in leucocyte adhesion; MILA; ss.

```

```

FH Key Location/Qualifiers
FT tata_signal 261..266
FT /tag= a
FT protein_bind 213..222
FT /tag= b
FT /number= 1
FT /note="NF-kappaB binding sequence"
FT protein_bind 227..236
FT /tag= c
FT /number= 2
FT /note="NF-kappaB binding sequence"
FT exon 303..472
FT /tag= d
FT /number= 2
FT /note="cDNA clone 41"
FT intron 473..1023
FT /tag= e
FT /number= 1
FT exon 1024..1032
FT /tag= f
FT /number= 2
FT WO9013300-A.
PN 15-NOV-1990.
PD 27-APR-1990; U02357.
PF 28-APR-1989; US-345151.
PR 01-JUN-1989; US-359516.
PR 18-DEC-1989; US-452675.
PA (BIOG-) BIOGEN INC.
PI Hession C, Lobb RR, Goelz SE, Born L, Benjamin CD;
PI Rosa MD;
DR WPI: 90-361248/48.
PT Endothelial cell adhesion mols. - MILAs and DNA encoding them and
PT inhibition-detection of binding of leukocytes to endothelial
PT cells
PS Disclosure; Fig 8; 136pp; English.
CC The 9 nucleotides N between bases 4013 and 4024 were not legible in
CC the specification.
CC VCAM1 clones were isolated by screening a human genomic EMBL3
CC library with a 32P-labeled 30 base oligomer probe homologous to the
CC 5' end of the VCAM1 cDNA.
CC NF-kappaB DNA binding activity is stimulated in endothelial cells
CC by IL-1 and TNF.
CC The promoter sequence is useful to construct vectors inducible by
CC cytokines, and bacterial LPS, or other agents found to induce
CC expression of ELAMs in endothelial cells. Such vectors may be useful
CC in gene transfer assays wherein the inducible promoter is positioned
CC so that it drives transcription of a reporter gene such as
CC chloramphenicol.
CC See also Q06686-91.
SQ Sequence 1032 BP; 273 A; 192 C; 216 G; 342 T;

Query Match 100.0%; Score 20; DB 1; Length 1032;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaaggaggattccctcc 20
Db 221 CCTGAAGGGATTCCCTCC 240

RESULT 5
Q90103
ID Q90103 standard; DNA; 19 BP.
AC Q90103;
DT 11-JAN-1996 (first entry)
DE VCAM-1 expression inhibiting oligonucleotide.
KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;
KW transcriptional regulatory factor; diagnosis; treatment; restenosis;
KW atherosclerosis; inflammatory disease; ds.
OS Synthetic.
PN WO9512415-A1.
PD 11-MAY-1995.

```

```

PF 07-NOV-1994; U12797.
PR 05-NOV-1993; US-147878.
PA (ISIS-) ISIS PHARM INC.
PI (UYEM-) UNIV EMORY.
PI Bennett CF, Medford RM;
DR WPI; 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17: Page 34; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 19 BP; 3 A; 5 C; 5 G; 6 T;

Query Match 90.0%; Score 18; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ttgaaggattccctcc 20
Db 1 TTGAAGGATTTCCTCC 18

RESULT 6
N80038
ID N80038 standard; cDNA; 1548 BP.
AC N80038;
DE 30-OCT-1990 (first entry)
DE Colony stimulating gene region and promoter region.
KW Colony stimulating factor; promoter.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_feature 796..1315
FT /*tag= a
FT /note="claimed CSF-1 gene promoter region"
PN D83808213-A.
PD 22-SEP-1988.
PR 11-MAY-1988; 808213.
PR 13-MAR-1987; JP-056890.
PA (GREG) Green Cross Corp.
PI Murakami K, Nakakubo H, Kaneda T, Nagai M, Arimura H;
DR WPI; 88-272091/39.
PT New promoter sequence of colony stimulating factor gene -
PT for controlling expression of protein in mammalian cells.
PS Disclosure; d claim 1, page 2; 12pp; german.
CC The colony stimulating factor-1 gene promoter region is useful for
CC controlling expression of urokinase, hepatitis B antigen, human serum
CC albumin and interferons in, eg mouse L cells or cell lines derived from
CC T cells and tumours.
SQ Sequence 1548 BP; 349 A; 388 C; 498 G; 313 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1548;
Best Local Similarity 89.5%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cttgaaggattccctcc 20
Db 862 CTTCAAGGATTTCCTCC 880

RESULT 7
N80398
ID N80398 standard; DNA; 1549 BP.
AC N80398;
DE 11-NOV-1990 (first entry)
DE Region upstream of colony stimulating factor-1 gene.
KW Colony stimulating factor-1; promoter region.

OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_feature 797..1316
FT /*tag= a
FT /note="claimed promoter sequence"
FT misc_feature 1317..
FT /*tag= b
FT /note="known sequence"
FT misc_feature 1193..1240
FT /*tag= c
FT /note="G-T cluster"
FT misc_feature 1063..1071
FT /*tag= d
FT /note="region homologous to enhancer-core region of
FT immunoglobulin heavy chain gene"
FT misc_feature 1186..1194
FT /*tag= e
FT /note="region homologous to enhancer-core region of
FT immunoglobulin heavy chain gene"
FT misc_feature 797..806
FT /*tag= f
FT /note="region homologous to consensus sequence upstream
FT from lymphokine gene"
FT misc_feature 869..878
FT /*tag= g
FT /note="region homologous to consensus sequence upstream
FT from lymphokine gene"
FT misc_feature 1257..1266
FT /*tag= h
FT /note="region homologous to consensus sequence upstream
FT from lymphokine gene"
FT cds 1495..
FT /*tag= i
FT intron 1534..
FT /*tag= j
PN GB2204042-A.
PD 02-NOV-1988.
PR 11-MAR-1988; 005799.
PR 13-MAR-1987; JP-056890.
PA (GREG) Green Cross Corp.
PI Murakami K, Nakakubo H, Kaneda T, Nagai M, Arimura H;
DR WPI; 88-272091/39.
PT New promoter sequence of colony stimulating factor gene -
PT for controlling expression of protein in mammalian cells.
PS Disclosure; pp; English.
CC The DNA contains a claimed CSF-1 promoter region, which is useful for
CC controlling expression of urokinase, hepatitis B antigen, human serum
CC albumin and interferons, etc. in mouse L cells or cell lines derived from
CC T cells and tumours.
SQ Sequence 1549 BP; 349 A; 388 C; 499 G; 313 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1549;
Best Local Similarity 89.5%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cttgaaggattccctcc 20
Db 863 CTTCAAGGATTTCCTCC 881

RESULT 8
T67638/c
ID T67638 standard; DNA; 378 BP.
AC T67638;
DT 11-JUL-1997 (first entry)
DE H. pylori flagella-associated protein ORF 3942217.aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis; ds.
OS Helicobacter pylori.
FH Key Location/Qualifiers

```

```

FT cds      1. .291
FT          /tag- a
FT          /transl_except= (pos: 223. .225, aa: Xaa)
FT          /note= "Xaa = unknown, no stop codon given"
PN WO9640893-Al.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI; 97-052306/05.
DR P-PSDB; W20473.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 1; Page 278; 1481pp; English.
CC This sequence encodes a H. pylori flagella-associated protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 378 BP; 121 A; 59 C; 94 G; 103 T;

Query Match      77.0%; Score 15.4; DB 1; Length 378;
Best Local Similarity 94.1%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 tgaaggatttcctcc 20
      ||| ||||| ||||| |||
DB      366 TGAATGGATTTCCTCC 350

RESULT 9
ID T68240/G
AC T68240;
DE 21-JUL-1997 (first entry)
DE H. pylori flagella-associated protein ORF hpell1122orf5.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis; ds.
OS Helicobacter pylori.
FH Key      Location/Qualifiers
FT cds      1. .387
FT          /tag- a
FT          /note= "no stop codon given"
PN WO9640893-Al.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI; 97-052306/05.
DR P-PSDB; W20987.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 1; Page 984; 1481pp; English.
CC The present sequence encodes a H. pylori flagella-associated protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori

infection or to identify H. pylori polypeptide binding compounds,
The genomic sequence of H. pylori (ATCC 55679) was determined from
overlapping contigs generated by mechanically shearing the bacterial
DNA. The sequences were analysed for ORF of at least 180 nucleotides,
and the predicted coding regions defined by computer evaluation. To
identify likely H. pylori antigens for vaccine development, the amino
acid sequences predicted from various ORF were analysed for significant
homology to other known or exported membrane proteins. Having identified
and determined the sequences of interest, particular regions can be
isolated from H. pylori by PCR amplification for recombinant polypeptide
production, e.g. in E. coli hosts.
Sequence 387 BP; 126 A; 60 C; 96 G; 105 T;

Query Match      77.0%; Score 15.4; DB 1; Length 387;
Best Local Similarity 94.1%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 tgaaggatttcctcc 20
      ||| ||||| ||||| |||
DB      375 TGAATGGATTTCCTCC 359

RESULT 10
ID V86164
ID V86164 standard; CDNA; 477 BP.
AC V86164;
DE 27-APR-1999 (first entry)
DE EST clone I5.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN WO9845435-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06954.
PR 10-APR-1997; US-935913.
PR (GENY ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070076/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 142-143; 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC therapy. The EST sequences are also stated to be useful for gene
SQ Sequence 477 BP; 90 A; 136 C; 144 G; 105 T;

Query Match      77.0%; Score 15.4; DB 1; Length 477;
Best Local Similarity 94.1%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 tgaaggatttcctcc 20
      ||| ||||| ||||| |||
DB      38 TGAATGGATTTCCTCC 54

```

```

RESULT 11
ID T18011/c
AC T18011 standard; cDNA; 1629 BP.
AT 03-OCT-1996 (first entry)
DE Beta-1onone 4-methylene gp. to keto gp. converting enzyme cDNA.
KW Converting enzyme; 4-methylene group; beta-1onone compound;
KW keto group; microbial host cell; transformation; recombinant;
KW vector; production; astaxanthin; 4-ketoxanthin; canthaxanthin;
KW echinenone; ketocarotenoids; NIES-144; biosynthesis; ds.
OS Haematococcus pluvialis.
FH Key Location/Qualifiers
FT cds 168..1082
FT /*tag= a
PN WO9606172-A1.
PD 29-FEB-1996.
PF 18-AUG-1995; J01640.
PR 23-AUG-1994; JP-198775.
PR 19-SEP-1994; JP-223798.
PR 07-MAR-1995; JP-047266.
PA (KIRI ) KIRIN BEER KK.
PI Kajiwa S, Kondo K, Misawa N;
DR WPI; 96-151375/15.
DR P-PSDB; R92096.
PT Polypeptide(s) and their DNA which introduce a keto-gp. in
PT beta-1onone cpds. - for prodn. of keto:carotenoid(s) in
PT transformed E. coli.
PS Claim 10; Pages 33-36; 63pp; Japanese.
CC The present sequence encodes a converting enzyme, which converts
CC the 4-methylene gp. of a beta-1onone cpd. into a keto gp.
CC Microbial host cells (pref. E. coli) transformed with a recombinant
CC vector contg. the cDNA, can be used for the prodn. of astaxanthin,
CC 4-ketoxanthin, canthaxanthin, echinenone and other
CC ketocarotenoids. The cDNA was isolated from a cDNA expression
CC library obtd. from Haematococcus pluvialis NIES-144, in which
CC astaxanthin biosynthesis was induced by 45 mM acetic acid in
CC the culture medium.
CC Sequence 1629 BP; 348 A; 458 C; 461 G; 362 T;
SQ

Query Match 77.0%; Score 15.4; DB 1; Length 1629;
Best Local Similarity 94.1%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 cttgaaggatttcct 18
Db 1465 CTCGAGGGATTTCCT 1449

RESULT 12
ID T22912/c
AC T22912 standard; cDNA to mRNA; 316 BP.
AT 29-AUG-1996 (first entry)
DE Human gene signature HUMGS04637.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1243; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.

```

---

```

CC double-stranded DNA) which comprises one of the 7837 "Gs" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared from
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 316 BP; 110 A; 45 C; 54 G; 92 T;

Query Match 76.0%; Score 15.2; DB 1; Length 316;
Best Local Similarity 85.0%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggatttcctcc 20
Db 46 CCTTGAAGGATTTCCTCC 27

RESULT 13
ID V52170 standard; DNA; 6171 BP.
AC V52170;
AT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:37.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI; 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 371-374; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 6171 BP; 1645 A; 1424 C; 1188 G; 1914 T;

```

```

Query Match      76.0%; Score 15.2; DB 1; Length 6171;
Best Local Similarity 85.0%; Pred. No. 59;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccttgaaggatttcctcc 20
    ||||| | |||||
Db 4682 CCTTGACCGAATTCCTCC 4701

RESULT 14
Q90105
ID T39170 standard; DNA; 34 BP.
AC Q90105;
DT 11-JAN-1996 (first entry)
DE VCAM-1 expression inhibiting oligonucleotide.
KW Vascular cell adhesion molecule; VCAM-1; inhibitory oligonucleotide;
KW transcriptional regulatory factor; diagnosis; treatment; restenosis;
KW atherosclerosis; inflammatory disease; ds.
OS Synthetic.
PN WO9512415-A1.
PD 11-MAY-1995.
PF 07-NOV-1994; U12797.
PR 05-NOV-1993; US-147878.
PA (ISIS-) ISIS PHARM INC.
PA (UYEM-) UNIV EMORY.
PI Bennett CF, Medford RM;
DR WPI: 95-193802/25.
PT Oligo-nucleotide(s) which modulate vascular cell adhesion molecule
PT expression by binding a transcription regulatory element - used to
PT diagnose and treat atherosclerosis, restenosis or inflammatory
PT disease
PS Claim 17; Page 35; 49pp; English.
CC Q90100-Q90111 bind the vascular cell adhesion molecule (VCAM)-1
CC gene transcriptional regulatory factor, therefore inhibiting the
CC expression of VCAM-1. They can be used in the diagnosis and
CC treatment of restenosis, atherosclerosis and inflammatory
CC diseases.
SQ Sequence 34 BP; 7 A; 9 C; 7 G; 11 T;

Query Match      75.0%; Score 15; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 gaaggatttcctc 19
    ||||| |||||
Db 1 GAAGGGATTTCCTC 15

RESULT 15
T39170
ID T39170 standard; DNA; 1230 BP.
AC T39170;
DT 04-MAR-1997 (first entry)
DE Bacillus stearothermophilus amino acid amidohydrolase DNA.
KW Amino acid amidohydrolase; carbamoylase; L-methionine; stereospecific;
KW thermostable; N-carbamoyl; cam gene; ds.
OS Bacillus stearothermophilus.
FH Key Location/Qualifiers
FT cds
FT 1..1230
FT /*tag= a
FT /product= amino acid amidohydrolase
FT /transl_except= pos:1..3, aa:Met
FT /note= "TTG initiation codon"

PN FR2728905-A1.
PD 05-JUL-1996.
PF 29-DEC-1994; 015838.
PR 29-DEC-1994; FR-015838.
PA (RHON) RHONE-POULENC NUTRITION ANIMALE.
PI Batisse N, Hallett JN, Lecocq FM, Sakanyan V;
PI Weigel P;
DR WPI: 96-335881/34.
DR P-PSDB; W03544.

```

---

```

PT New stereospecific, heat-stable amino acid amidohydrolase - from
PT B.stearothermophilus and related DNA, esp. for prodn. of L-Met.
PS Claim 4; Page 9-11; 17pp; French.
CC PstI fragments of Bacillus stearothermophilus NCIB 8224 genomic DNA
CC were inserted into pBR322 and used for transforming E.coli hosts.
CC Selection was on medium which contained tetracycline but lacked
CC arginine. All positive transformants carried plasmids with a 4.7 kb
CC insert. The present sequence is a fragment of the insert. Amino
CC acid amidohydrolase encoded by the fragment is stereospecific and
CC heat-stable (maximum activity at 55-60 deg.C). The enzyme
CC hydrolyses N-carbamoyl amino acid derivs. to L-amino acids and is
CC particularly useful for production of L-methionine. 237 T;
SQ Sequence 1230 BP; 268 A; 312 C; 413 G; 237 T;

Query Match      75.0%; Score 15; DB 1; Length 1230;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 gaaggatttcctc 19
    ||||| |||||
Db 460 GAAGGGATTTCCTC 474

Search completed: May 23, 2000, 09:42:02
Job time: 8834 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 07:12:13 ; Search time 1428.99 seconds  
(without alignments)  
56.729 Million cell updates/sec

Title: US-08-945-805-1  
Perfect score: 20  
Sequence: 1 ccttgaaggattccctcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: em\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query
--------	---	-------

No.	Score	Match	Length	DB	ID	Description
C 1	17	85.0	350	22	R25866	R25866 yh42f12.r1
C 2	16.8	84.0	258	74	AW190931	AW190931 x166g08.x
C 3	16.8	84.0	265	45	AI367826	AI367826 qg40b08.x
C 4	16.8	84.0	371	38	AA775085	AA775085 ac76h01.s
C 5	16.8	84.0	376	27	AA055416	AA055416 zf18b08.r
C 6	16.8	84.0	426	22	H02051	H02051 y134b05.r1
C 7	16.8	84.0	436	21	T93077	T93077 y24d04.s1
C 8	16.8	84.0	466	27	AA054954	AA054954 zfi18b08.s
C 9	16.8	84.0	496	36	AA624941	AA624941 vni1e11.r
C 10	16.8	84.0	740	42	AI158037	AI158037 ud26b11.r
C 11	16.4	82.0	145	62	AI902043	AI902043 618014D02
C 12	16.4	82.0	146	60	AV135061	AV135061 AV135061
C 13	16.4	82.0	488	84	B43292	B43292 HS-1057-B1-
C 14	16	80.0	423	21	T67132	T67132 yf53a02.s4
C 15	15.8	79.0	340	43	AU032025	AU032025 AU032025
C 16	15.8	79.0	343	81	AA46474	AA46474 86035 MAR
C 17	15.8	79.0	350	26	X96620	X96620 HSKAP103B H
C 18	15.8	79.0	370	30	AA199538	AA199538 mt54e11.r
C 19	15.8	79.0	371	30	AA199543	AA199543 mt54g11.r
C 20	15.8	79.0	393	50	AV046928	AV046928 AV046928
C 21	15.8	79.0	394	105	AQ285175	AQ285175 RPCI11-79
C 22	15.8	79.0	401	21	T84265	T84265 yd31h06.r1
C 23	15.8	79.0	402	22	R78549	R78549 y173f02.s1
C 24	15.8	79.0	419	105	AQ339544	AQ339544 HS_5022_A
C 25	15.8	79.0	422	43	C98642	C98642 C98642 Rice
C 26	15.8	79.0	432	48	AI592638	AI592638 mt54e11.y
C 27	15.8	79.0	437	37	AA698073	AA698073 HL03733.5
C 28	15.8	79.0	439	21	T85814	T85814 yd56g10.r1
C 29	15.8	79.0	444	102	AQ115885	AQ115885 RPCI11-56
C 30	15.8	79.0	453	48	AI592642	AI592642 mt54g11.y
C 31	15.8	79.0	455	74	AW210361	AW210361 ul54a05.y
C 32	15.8	79.0	457	61	AI871501	AI871501 w167g10.x
C 33	15.8	79.0	462	60	AI806038	AI806038 te48a04.x
C 34	15.8	79.0	467	21	R18596	R18596 yf96c12.r1
C 35	15.8	79.0	468	27	AA026534	AA026534 zj99c06.r
C 36	15.8	79.0	483	29	AA144290	AA144290 mr16f12.r
C 37	15.8	79.0	497	60	AI794385	AI794385 fc43g09.y
C 38	15.8	79.0	514	48	AI562650	AI562650 TENS2637
C 39	15.8	79.0	542	43	AI205150	AI205150 ao84c01.x
C 40	15.8	79.0	542	103	AQ227469	AQ227469 HS_2015_B
C 41	15.8	79.0	543	33	AA443169	AA443169 zx98d11.r
C 42	15.8	79.0	563	47	AI541900	AI541900 SD07859.5
C 43	15.8	79.0	666	102	AQ082547	AQ082547 RPCI11-50
C 44	15.8	79.0	675	46	AI406103	AI406103 GH26293.5
C 45	15.8	79.0	682	74	AW182810	AW182810 xj64b07.x

## ALIGNMENTS

RESULT 1  
 R25866/c  
 LOCUS yh42f12.r1 Soares placenta Nb2HP Homo sapiens cDNA clone  
 DEFINITION IMAGE:132431 5', mRNA sequence.  
 R25866  
 VERSION R25866.1 GI:782001  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 350)  
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,K., Williamson,A., Wohlmann,P. and Wilson,R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On Sep 21, 1992 this sequence version replaced gi:276321.

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Insert Size: 741  
 High quality sequence stops: 269  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 741 Std Error: 0.00  
 Seq primer: M13Rpl  
 High quality sequence stop: 269.

FEATURES  
 source  
 1..350  
 /organism="Homo sapiens"  
 /db\_xref="GDB:538033"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:132431"  
 /clone\_lib="Soares placenta Nb2HP"  
 /sex="Female"  
 /dev\_stage="placenta obtained at birth (full term)"  
 /lab\_host="PH10B (ampicillin resistant)"  
 /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACGAGAAATCGCGCCGAGAAATTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo. "

BASE COUNT 82 a 67 c 81 g 118 t 2 others  
 ORIGIN  
 Query Match 85.0%; Score 17; DB 22; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 17; Conservative 0;

QY 4 tgaagggtttccctcc 20  
 |||||  
 Db 233 TGAAGGATTCCTCC 217

RESULT 2  
 AW190931/c  
 LOCUS x166g08.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone  
 DEFINITION mRNA sequence.  
 ACCESSION AW190931  
 VERSION AW190931.1 GI:6465411  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 258)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189734.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Life Technologies catalog #: 11548-013  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

```

Seq primer: -40UP from Gibco
High quality sequence stop: 244.
Location/Qualifiers
1. .258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2679710"
/clone_lib="NCI_CGAP_Panl"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT      62 a  60 c  56 g  80 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 74; Length 258;
Best Local Similarity 90.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccttgaagggttcctcc 20
||||| ||||||| |||
Db 191 CCTTGAGGGGATTCCTTCC 172

RESULT 3
AI367826/c
LOCUS
DEFINITION q40508.x1 Soares.NHMPu.S1 Homo sapiens cDNA clone IMAGE:1934967
3', similar to gb:M60484.nral PROTEIN PHOSPHATASE PP2A-BETA,
CATALYTIC SUBUNIT (HUMAN);, mRNA sequence.
ACCESSION AI367826
VERSION AI367826.1 GI:4137571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 265)
AUTHORS Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT Unpublished (1997)
On Jan 17, 1998 this sequence version replaced gi:1900238.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 1908 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. .265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1934967"
/clone_lib="Soares.NHMPu.S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7b-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBHH19) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools

```

```

consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      108 a  59 c  38 g  60 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 45; Length 265;
Best Local Similarity 90.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccttgaagggttcctcc 20
||||| ||||||| |||
Db 101 CTTTGAAGGGATTCCTTCC 82

RESULT 4
AA775085/c
LOCUS
DEFINITION ac76h01.s1 Stratagene lung (#937210) Homo sapiens cDNA clone
IMAGE:868561 3', mRNA sequence.
ACCESSION AA775085
VERSION AA775085.1 GI:2834419
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 371)
AUTHORS Hallier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900223.
Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 362.
Location/Qualifiers
1. .371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:868561"
/clone_lib="Stratagene lung (#937210)"
/sex="male"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. normal lung. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      95 a  82 c  93 g  101 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 38; Length 371;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccttgaagggttcctcc 20
||||| ||||||| |||
Db 184 CCTTGAGGGGATTCCTTCC 165

RESULT 5

```

```

AA055416/c
LOCUS       AA055416       376 bp      mRNA          EST          01-FEB-1997
DEFINITION zfl8b08.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:377271 5', mRNA sequence.
ACCESSION  AA055416
VERSION     AA055416.1  GI:1547754
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 376)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevisakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE     Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
COMMENT   On Jan 24, 1995 this sequence version replaced gi:634307.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1069 Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 365.
FEATURES   source
            Location/Qualifiers
                1..376
                    /organism="Homo sapiens"
                    /db_xref="GDB:1285527"
                    /db_xref="taxon:9606"
                    /clone_lib="Soares_fetal_heart_NbHH19W"
                    /sex="unknown"
                    /dev_stage="19 weeks"
                    /lab_host="PH108 (ampicillin resistant)"
                    /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                    strand cDNA was primed with a Not I - oligo(dT) primer [5',
                    TGTTCACATCTGAAGTGGGAGCGCGCATCTTTTTTTTTTTTTTTT 3'],
                    double-stranded cDNA was size selected, ligated to Eco RI
                    adapters (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of a modified pT7T3 vector
                    (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by
                    M.Fatima Bonaldo. This library was constructed from the
                    same fetus as the fetal lung library, Soares fetal lung
                    NBHL19W."
BASE COUNT  109 a 70 c 95 g 102 t
ORIGIN
            Query Match      84.0%; Score 16.8; DB 27; Length 376;
            Best Local Similarity 90.0%; Pred. No. 1.6e+02;
            Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctcc 20
      ||||| ||||| |||
Db 130 CCTTGAGGGGATTTCCTTC 111

RESULT 6
H02051
LOCUS     H02051       426 bp      mRNA          EST          20-JUN-1995
DEFINITION yj34b05.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:118663 3', mRNA sequence.
ACCESSION  H02051
VERSION     H02051
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 426)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
            The WashU-Merck EST Project
            Unpublished (1995)
            On May 9, 1995 this sequence version replaced gi:803314.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 697
            High quality sequence stops: 337
            Source: IMAGE Consortium, LNL
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 697 Std Error: 0.00
            Seq primer: M13Rpl
            High quality sequence stop: 337.
FEATURES   source
            Location/Qualifiers
                1..426
                    /organism="Homo sapiens"
                    /db_xref="GDB:562614"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:150609"
                    /clone_lib="Soares placenta Nb2HP"
                    /sex="Female"
                    /dev_stage="placenta obtained at birth (full term)"
                    /lab_host="PH108 (ampicillin resistant)"
                    /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                    strand cDNA was primed with a Not I - oligo(dT) primer [5',
                    AACTGGAGATTCGCGCGCGAGGAATTTTTTTTTTTTTTTT 3'],
                    double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization. Library
                    constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT  129 a 82 c 105 g 109 t 1 others
ORIGIN
            Query Match      84.0%; Score 16.8; DB 22; Length 426;
            Best Local Similarity 90.0%; Pred. No. 1.6e+02;
            Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctcc 20
      ||||| ||||| |||
Db 134 CCTTGAGGGGATTTCCTTC 153

RESULT 7
T93077/c
LOCUS     T93077       436 bp      mRNA          EST          22-MAR-1995
DEFINITION ye24d04.s1 Stragatene lung (#937210) Homo sapiens cDNA clone
IMAGE:118663 3', mRNA sequence.
ACCESSION  T93077
VERSION     T93077.1  GI:724990
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 436)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 23  
Source: IMAGE Consortium, LLNL This clone is available royalty-free  
through LLNL; contact the IMAGE Consortium (info@image.llnl.gov)  
for further information.  
Seq primer: -21ml3  
High quality sequence stop: 437.

FEATURES  
source  
1. 436  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:486952"  
/db\_xref="taxon:9606"  
/clones="IMAGE:118663"  
/clone\_lib="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. normal lung. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 121 a 90 c 98 g 124 t 3 others  
ORIGIN  
Query Match 84.0%; Score 16.8; DB 21; Length 436;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ccttgaaggatttcctcc 20  
||||| ||||||| |||  
Db 184 CCTTGAGGGGATTCCTTCC 165

RESULT 8  
AA054954  
LOCUS  
DEFINITION zfl18b08.s1 Soares fetal heart NbHH19W Homo sapiens CDNA clone  
IMAGE:377271 3', mRNA sequence. EST 01-FEB-1997  
AA054954  
ACCESSION  
VERSION AA054954.1 GI:1547320  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 466)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 436)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 23  
Source: IMAGE Consortium, LLNL This clone is available royalty-free  
through LLNL; contact the IMAGE Consortium (info@image.llnl.gov)  
for further information.  
Seq primer: -21ml3  
High quality sequence stop: 437.

FEATURES  
source  
1. 436  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:486952"  
/db\_xref="taxon:9606"  
/clones="IMAGE:118663"  
/clone\_lib="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. normal lung. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 121 a 90 c 98 g 124 t 3 others  
ORIGIN  
Query Match 84.0%; Score 16.8; DB 21; Length 436;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ccttgaaggatttcctcc 20  
||||| ||||||| |||  
Db 184 CCTTGAGGGGATTCCTTCC 165

RESULT 8  
AA054954  
LOCUS  
DEFINITION zfl18b08.s1 Soares fetal heart NbHH19W Homo sapiens CDNA clone  
IMAGE:377271 3', mRNA sequence. EST 01-FEB-1997  
AA054954  
ACCESSION  
VERSION AA054954.1 GI:1547320  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 466)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On Nov 29, 1993 this sequence version replaced gi:430144.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1069 Std Error: 0.00  
Seq primer: -40M13 fwd. from Amersham  
High quality sequence stop: 356.

FEATURES  
source  
1. 466  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:1285527"  
/db\_xref="taxon:9606"  
/clones="IMAGE:377271"  
/clone\_lib="Soares\_fetal\_heart\_NbHH19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGTGGAGCGCCGCACTCTTTTCTTTTCTTTT 3',  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NbHH19W."

BASE COUNT 126 a 97 c 76 g 164 t 3 others  
ORIGIN  
Query Match 84.0%; Score 16.8; DB 27; Length 466;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ccttgaaggatttcctcc 20  
||||| ||||||| |||  
Db 393 CCTTGAGGGGATTCCTTCC 412

RESULT 9  
AA624941/c  
LOCUS  
DEFINITION AA624941 496 bp mRNA EST 14-OCT-1997  
vn91e11.r1 Stratagene mouse heart (#937316) Mus musculus CDNA clone  
IMAGE:1039340 5' similar to gb:M95106 Mus musculus CAMP response  
element binding protein (MOUSE);, mRNA sequence.  
AA624941  
ACCESSION  
VERSION AA624941.1 GI:2528817  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 496)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
On May 9, 1995 this sequence version replaced gi:802286.  
Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:579364  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 446.

#### FEATURES

source

```
1. .496
Location/Qualifiers
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:1039340"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
BASE COUNT 150 a 123 c 122 g 101 t
ORIGIN
```

Query Match 84.0%; Score 16.8; DB 36; Length 496;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 ccttgaggaggttcctcc 20
||||| |||||||
Db 445 CCTGAAGGATTCCTTC 426
```

RESULT 10  
AII58037/c  
LOCUS  
DEFINITION  
AII58037 740 bp mRNA EST 30-SEP-1998  
ud26b11.r1 Soares\_thymus\_2NbMT Mus musculus cDNA clone  
IMAGE:1447005 5' similar to gb:M34356 CAMP-RESPONSE ELEMENT BINDING  
PROTEIN (HUMAN); gb:M95106 Mus musculus CAMP response element  
binding protein (MOUSE);, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AII58037.1 GI:3686506  
house mouse.  
Mus musculus

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 740)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
On Jan 19, 1998 this sequence version replaced gi:2153477.  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:920321  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 510.  
Location/Qualifiers  
1. .740  
source

#### FEATURES

source

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1447005"
/clone_lib="Soares_thymus_2NbMT"
/sex="male"
/tissue_type="thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGTGGACGCGCGGTGTGTGTGTGTGTGTGTGTGT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through two  
rounds of normalization, and was constructed by Bento  
Soares and M.Fatima Bonaudo."
```

BASE COUNT 220 a 185 c 174 g 154 t 7 others  
ORIGIN

Query Match 84.0%; Score 16.8; DB 42; Length 740;  
Best Local Similarity 90.0%; Pred. No. 1.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 ccttgaggaggttcctcc 20
||||| |||||||
Db 369 CCTGAAGGATTCCTTC 350
```

RESULT 11  
AII58037/c

LOCUS  
DEFINITION  
AII58037 145 bp mRNA EST 27-JUL-1999  
618014D02.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AII58037.1 GI:5608466  
EST.  
Zea mays.  
Zea mays

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
Poaceae; Zea.  
1 (bases 1 to 145)  
Walbot, V.  
Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
On Mar 10, 1998 this sequence version replaced gi:2948871.

CONTACT  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 618014 row: D column: 02.  
Location/Qualifiers  
1. .145  
/organism="Zea mays"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/clone\_lib="618 - Inbred Tassel cDNA Library"  
/tissue\_type="tassel"  
/dev\_stage="tassel length from 0.1 to 2.5 cm"  
/lab\_host="XLOLR"

#### FEATURES

source

```

/notes="Organ: tassal; Vector: PAD-GAL4-2.1 (Hybridap);
Inbred tassal library from Schmidt lab"
BASE COUNT      40 a 46 c 29 g 30 t
ORIGIN

Query Match      82.0%; Score 16.4; DB 62; Length 145;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttgaaggattccctc 20
|||||
Db 58 TTGAAGGATTTCCTCC 41

RESULT 12
AV135061
LOCUS
DEFINITION AV135061 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
clone 2810013C11, mRNA sequence.
ACCESSION AV135061
VERSION AV135061.1 GI:5320791
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 146)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Alzawa,K.,
Akahira,S., Akiyama,S., Fukuda,S., Fukunishi,Y., Funayama,T.,
Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
On May 9, 1996 this sequence version replaced gi:1301786.
JOURNAL Genome Science Laboratory
COMMENT Contact: Chie Owa
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@tc.riken.go.jp
Thermotabilization and thermostabilization of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..146
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2810013C11"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"
BASE COUNT      34 a 22 c 30 g 60 t
ORIGIN

Query Match      82.0%; Score 16.4; DB 60; Length 146;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cttgaaggattccctc 19
|||||
Db 28 CTTTAAGGATTTCCTCC 45

/notes="Organ: tassal; Vector: PAD-GAL4-2.1 (Hybridap);
Inbred tassal library from Schmidt lab"
BASE COUNT      40 a 46 c 29 g 30 t
ORIGIN

Query Match      82.0%; Score 16.4; DB 62; Length 145;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttgaaggattccctc 20
|||||
Db 58 TTGAAGGATTTCCTCC 41

RESULT 12
AV135061
LOCUS
DEFINITION AV135061 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
clone 2810013C11, mRNA sequence.
ACCESSION AV135061
VERSION AV135061.1 GI:5320791
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 146)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Alzawa,K.,
Akahira,S., Akiyama,S., Fukuda,S., Fukunishi,Y., Funayama,T.,
Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
On May 9, 1996 this sequence version replaced gi:1301786.
JOURNAL Genome Science Laboratory
COMMENT Contact: Chie Owa
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@tc.riken.go.jp
Thermotabilization and thermostabilization of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..146
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2810013C11"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"
BASE COUNT      34 a 22 c 30 g 60 t
ORIGIN

Query Match      82.0%; Score 16.4; DB 60; Length 146;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cttgaaggattccctc 19
|||||
Db 28 CTTTAAGGATTTCCTCC 45

/notes="Organ: tassal; Vector: PAD-GAL4-2.1 (Hybridap);
Inbred tassal library from Schmidt lab"
BASE COUNT      40 a 46 c 29 g 30 t
ORIGIN

Query Match      82.0%; Score 16.4; DB 62; Length 145;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttgaaggattccctc 20
|||||
Db 58 TTGAAGGATTTCCTCC 41

RESULT 13
B43292
LOCUS
DEFINITION B43292 HS-1057-B1-A06-MR.abi CIT Human Genomic Sperm Library C Homo
sapiens genomic clone Plate=CT 779 Col=11 Row=B, genomic survey
sequence.
ACCESSION B43292
VERSION B43292.1 GI:2548126
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 488)
AUTHORS Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
CONSTRUCTION OF A CHARACTERIZED CLONE RESOURCE FOR GENOMIC
SEQUENCING: Generation and Preliminary Analysis of 20,000 Sequence
Tagged Connectors
Unpublished (1997)
Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackron@u.washington.edu
Sequence Tagged Connector
Plate: CT 779 row: B column: 11
Class: BAC ends
High quality sequence stop: 488.
FEATURES
source
Location/Qualifiers
1..488
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 779 Col=11 Row=B"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      117 a 99 c 106 g 166 t
ORIGIN

Query Match      82.0%; Score 16.4; DB 84; Length 488;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cttgaaggattccctc 19
|||||
Db 82 CTTGAGGGATTTCCTC 99

RESULT 14
T67132
LOCUS
DEFINITION T67132 YAS3a02.s4 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:66602 3', mRNA sequence.
ACCESSION T67132
VERSION T67132.1 GI:676572
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 423)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisan,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
TITLE

```

JOURNAL  
COMMENT

Unpublished (1995)  
Other\_ESTs: ya53a02.r4.exp  
Contact: Willson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 401  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -21ml3  
High quality sequence stop: 401.

## FEATURES

source

1..423  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:66602"  
/clone\_lib="Soares fetal liver spleen 1NFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
15' AACGCGAAGATTAATTAAGATCTTTTTTTTTTTTTTTT 3',  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT  
ORIGIN

95 a 103 c 113 g 90 t 22 others

Query Match

Best Local Similarity 80.0%; Score 16; DB 21; Length 423;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttgaagggtttcc 16

|||||

Db 307 CCTTGAAGGGATTCC 322

## RESULT 15

AU032025

LOCUS

AU032025 340 bp mRNA EST 20-OCT-1998

DEFINITION AU032025 Rice root Oryza sativa cDNA clone R3228\_62, mRNA sequence.

ACCESSION AU032025

VERSION AU032025.1 GI:3767998

KEYWORDS

EST.

SOURCE

Oryza sativa.

ORGANISM

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; Oryza. 340)

1 (bases 1 to 340)

Minobe.Y. and Sasaki.T.

Rice cDNA from root

Unpublished (1995)

JOURNAL

COMMENT

On Jan 14, 1998 this sequence version replaced gi:1798016.

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai, Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@abr.affrc.go.jp

PROJECT "RGP"

Location/Qualifiers

source

1..340  
/organism="Oryza sativa"  
/strain="Nipponbare, sub\_species Japonica"  
/db\_xref="taxon:4530"  
/clone="R3228\_62"  
/clone\_lib="Rice root"  
/note="Prepared from seedling root."

BASE COUNT  
ORIGIN

94 a 72 c 81 g 89 t 4 others

Query Match

Best Local Similarity 79.0%; Score 15.8; DB 43; Length 340;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttgaagggtttccctcc 20

|||||

Db 262 CTTAATGGGATTCCCTCC 280

Search completed: May 23, 2000, 09:21:39

Job time: 7766 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 07:14:21 ; Search time 71.7 Seconds  
(without alignments)  
36.258 Million cell updates/sec

Title: US-08-945-805-1

Perfect score: 20

Sequence: 1 ccttgaaggatttcctcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	15.4	77.0	1677	4	US-08-532-434-1
c 2	15.2	76.0	2475	4	US-07-684-965-5
c 3	14.4	72.0	2313	2	US-08-232-538-5
4	14.4	72.0	2352	2	US-08-232-538-17
5	14.4	72.0	2352	3	US-08-786-164-17
6	14.4	72.0	2651	3	US-08-786-164-5
7	14.2	71.0	21	3	US-08-410-654B-34
8	14.2	71.0	21	3	US-08-474-851-34
9	14.2	71.0	21	3	US-08-481-560-34
10	14.2	71.0	36	3	US-08-412-376-31
11	14.2	71.0	525	1	US-08-434-411-1
12	14.2	71.0	525	2	US-08-434-402-1
13	14.2	71.0	525	2	US-08-783-288-1
14	14.2	71.0	525	4	US-08-890-640-1
15	14.2	71.0	525	6	PCT-US94-12873-1
16	14.2	71.0	525	7	5194592-25
17	14.2	71.0	537	6	PCT-US94-12873-16
18	14.2	71.0	541	3	US-08-797-689-13
19	14.2	71.0	546	5	US-08-469-318-177
20	14.2	71.0	546	5	US-08-469-318-178
21	14.2	71.0	546	5	US-08-469-318-179
22	14.2	71.0	546	6	PCT-US95-01185-177
23	14.2	71.0	546	6	PCT-US95-01185-178
24	14.2	71.0	546	6	PCT-US95-01185-179
25	14.2	71.0	555	6	PCT-US94-12873-13
c 26	14.2	71.0	713	3	US-08-613-942-1
27	14.2	71.0	906	5	US-08-469-318-53

Sequence 53, Appl  
Sequence 56, Appl  
Sequence 59, Appl  
Sequence 60, Appl  
Sequence 72, Appl  
Sequence 75, Appl  
Sequence 78, Appl  
Sequence 80, Appl  
Sequence 84, Appl  
Sequence 85, Appl  
Sequence 87, Appl  
Sequence 56, Appl  
Sequence 59, Appl  
Sequence 60, Appl  
Sequence 72, Appl  
Sequence 75, Appl  
Sequence 78, Appl  
Sequence 80, Appl  
Sequence 84, Appl  
Sequence 85, Appl  
Sequence 87, Appl  
Sequence 56, Appl  
Sequence 59, Appl  
Sequence 60, Appl  
Sequence 72, Appl  
Sequence 75, Appl  
Sequence 78, Appl  
Sequence 80, Appl  
Sequence 84, Appl  
Sequence 85, Appl  
Sequence 87, Appl

ALIGNMENTS

RESULT 1

US-08-632-434-1/c

; Sequence 1, Application US/08632434

; Patent No. 5910433

; GENERAL INFORMATION:

; APPLICANT: KAJIWARA, Susumu

; APPLICANT: MISAHA, No. 5910433hiko

; TITLE OF INVENTION: KETO GROUP-INTRODUCING ENZYME, DNA

; TITLE OF INVENTION: KETO GROUP-INTRODUCING ENZYME, DNA

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/632,434

; FILING DATE: 23-APR-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 95/047266

; FILING DATE: 07-MAR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 94/223798

; FILING DATE: 19-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 94/198775

; FILING DATE: 23-AUG-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 81356/106

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1677 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

FEATURE:  
NAME/KEY: CDS  
LOCATION: 168..1127  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 168..1127  
US-08-632-434-1

Query Match 77.0%; Score 15.4; DB 4; Length 1677;  
Best Local Similarity 94.1%; Pred. No. 15;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 cttgaaggattccct 18  
|| |||||  
Db 1513 CTCGAAGGATTCCTC 1497

RESULT 2  
US-07-684-965-5/C  
Sequence 5, Application US/07684965  
Patent No. 5919649  
GENERAL INFORMATION:  
APPLICANT: HABENER, JOEL F.  
TITLE OF INVENTION: A-CAMP RESPONSIVE TRANSCRIPTIONAL  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESS: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue, N.W., Suite 300  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/684.965  
FILING DATE: 19910522

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0609.2470004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)833-7533  
TELEFAX: (202)833-8716

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2475 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 126..1106  
US-07-684-965-5

Query Match 76.0%; Score 15.2; DB 4; Length 2475;  
Best Local Similarity 85.0%; Pred. No. 21;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaaggattccctcc 20  
|||||  
Db 473 CCTGAAGAATTCCCTC 454

RESULT 3  
US-08-232-538-5  
Sequence 5, Application US/08232538  
Patent No. 5712380  
GENERAL INFORMATION:  
APPLICANT: Thomas, Kenneth A.  
TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL CELL  
GROWTH FACTOR  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: Merck & Co., Inc.  
STREET: P.O. Box 2000 126 E Lincoln Avenue  
CITY: Rahway  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065-0907

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.538  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen, John W.III  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 188881A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3905  
TELEFAX: (908) 594-4720

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2313 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-232-538-5

Query Match 72.0%; Score 14.4; DB 2; Length 2313;  
Best Local Similarity 93.8%; Pred. No. 56;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tgaaggattccctc 19  
|||||  
Db 1325 TCAGGCATTCCTC 1340

RESULT 4  
US-08-232-538-17  
Sequence 17, Application US/08232538  
Patent No. 5712380  
GENERAL INFORMATION:  
APPLICANT: Thomas, Kenneth A.  
TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL CELL  
GROWTH FACTOR  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: Merck & Co., Inc.  
STREET: P.O. Box 2000 126 E Lincoln Avenue  
CITY: Rahway  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065-0907

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/232,538  
;; FILING DATE:  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Wallen, John W.III  
;; REGISTRATION NUMBER: 35,403  
;; REFERENCE/DOCKET NUMBER: 188881A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (908) 594-3905  
;; TELEFAX: (908) 594-4720  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2352 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
US-08-232-538-17

Query Match 72.0%; Score 14.4; DB 2; Length 2352;  
Best Local Similarity 93.8%; Pred. No. 56;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 tgaagggtttccctc 19  
||||| |||||||  
Db 1085 TGAAGGCATTTCCCTC 1100

RESULT 5  
US-08-786-164-17  
; Sequence 17, Application US/08786164  
; Patent No. 5861484  
; GENERAL INFORMATION:  
; APPLICANT: THOMAS, KENNETH A.  
; APPLICANT: KENDALL, RICHARD L  
; TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL  
; TITLE OF INVENTION: CELL GROWTH FACTOR  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merck & Co., Inc.  
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
; CITY: Rahway  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07065-0900  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Microsoft Word 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/786,164  
; FILING DATE: 21-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark Hand, J  
; REGISTRATION NUMBER: 36,545  
; REFERENCE/DOCKET NUMBER: 18888DA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-594-3905  
; TELEFAX: 908-594-4720  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2352 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

;; TOPOLOGY: linear  
;; MOLECULE TYPE: Genomic DNA  
US-08-786-164-17

Query Match 72.0%; Score 14.4; DB 3; Length 2352;  
Best Local Similarity 93.8%; Pred. No. 56;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 tgaagggtttccctc 19  
||||| |||||||  
Db 1085 TGAAGGCATTTCCCTC 1100

RESULT 6  
US-08-786-164-5  
; Sequence 5, Application US/08786164  
; Patent No. 5861484  
; GENERAL INFORMATION:  
; APPLICANT: THOMAS, KENNETH A.  
; APPLICANT: KENDALL, RICHARD L  
; TITLE OF INVENTION: INHIBITOR OF VASCULAR ENDOTHELIAL  
; TITLE OF INVENTION: CELL GROWTH FACTOR  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merck & Co., Inc.  
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
; CITY: Rahway  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07065-0900  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Microsoft Word 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/786,164  
; FILING DATE: 21-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark Hand, J  
; REGISTRATION NUMBER: 36,545  
; REFERENCE/DOCKET NUMBER: 18888DA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-594-3905  
; TELEFAX: 908-594-4720  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2651 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-786-164-5

Query Match 72.0%; Score 14.4; DB 3; Length 2651;  
Best Local Similarity 93.8%; Pred. No. 58;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 tgaagggtttccctc 19  
||||| |||||||  
Db 1325 TGAAGGCATTTCCCTC 1340

RESULT 7  
US-08-410-654B-34  
; Sequence 34, Application US/08410654B

```
; Patent No. 5833976
; GENERAL INFORMATION:
; APPLICANT: Rene de Waal Malefyt
; APPLICANT: Di-Hwei Hsu
; APPLICANT: Anne O'Garra
; APPLICANT: Hergen Spits
; TITLE OF INVENTION: Use of Interleukin-10 to Treat
; TITLE OF INVENTION: Septic Shock
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7.5.3
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,654B
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/229,854
; FILING DATE: 19-APR-1994
; APPLICATION NUMBER: US 07/926,853
; FILING DATE: 06-AUG-1992
; APPLICATION NUMBER: US 07/742,129
; FILING DATE: 06-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Foulke, Cynthia L.
; REGISTRATION NUMBER: 32,364
; REFERENCE/DOCKET NUMBER: DX0221KQ1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-298-2987
; TELEFAX: 908-298-5388
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (oligonucleotide)
; US-08-410-654B-34

Query Match 71.0%; Score 14.2; DB 3; Length 21;
Best Local Similarity 84.2%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccttgaaggattccctc 19
   ||| ||||| |||||
Db 3 CCTGGAAGGATCTCCCC 21

RESULT 8
US-08-474-851-34
; Sequence 34, Application US/08474851
; Patent No. 5837232
; GENERAL INFORMATION:
; APPLICANT: Rene de Waal Malefyt
; APPLICANT: Di-Hwei Hsu
; APPLICANT: Anne O'Garra
; APPLICANT: Hergen Spits
; TITLE OF INVENTION: Use of An Interleukin-10 Antagonist to Treat
; TITLE OF INVENTION: A B Cell Mediated Autoimmune Disorder
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: 2000 Galloping Hill Road
```

```
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7.5.3
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,851
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/410,654
; FILING DATE: 24-MAR-1995
; APPLICATION NUMBER: US 08/229,854
; FILING DATE: 19-APR-1994
; APPLICATION NUMBER: US 07/926,853
; FILING DATE: 06-AUG-1992
; APPLICATION NUMBER: US 07/742,129
; FILING DATE: 06-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Foulke, Cynthia L.
; REGISTRATION NUMBER: 32,364
; REFERENCE/DOCKET NUMBER: DX0221KQ1GD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-298-2987
; TELEFAX: 908-298-5388
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (oligonucleotide)
; US-08-474-851-34

Query Match 71.0%; Score 14.2; DB 3; Length 21;
Best Local Similarity 84.2%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccttgaaggattccctc 19
   ||| ||||| |||||
Db 3 CCTGGAAGGATCTCCCC 21

RESULT 9
US-08-481-560-34
; Sequence 34, Application US/08481560
; Patent No. 5837293
; GENERAL INFORMATION:
; APPLICANT: Rene de Waal Malefyt
; APPLICANT: Di-Hwei Hsu
; APPLICANT: Anne O'Garra
; APPLICANT: Hergen Spits
; TITLE OF INVENTION: Use of Interleukin-10 to Modulate
; TITLE OF INVENTION: Inflammation or T-Cell Mediated
; TITLE OF INVENTION: Immune Function
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7.5.3
; SOFTWARE: Microsoft Word 6.0
```

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,560  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/410,654  
; FILING DATE: 24-MAR-1995  
; APPLICATION NUMBER: US 08/229,854  
; FILING DATE: 19-APR-1994  
; APPLICATION NUMBER: US 07/926,853  
; FILING DATE: 06-AUG-1992  
; APPLICATION NUMBER: US 07/742,129  
; FILING DATE: 06-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Foulke, Cynthia L.  
; REGISTRATION NUMBER: 32,364  
; REFERENCE/DOCKET NUMBER: DX0221KQ1GC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-298-2987  
; TELEFAX: 908-298-5388  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (oligonucleotide)  
; US-08-481-560-34

Query Match 71.0%; Score 14.2; DB 3; Length 21;  
Best Local Similarity 84.2%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cctgaaggaggttcctc 19  
||| ||||| |||||  
Db 3 CCTGAAGGAGGTCTCCCC 21

RESULT 10  
US-08-412-376-31  
; Sequence 31, Application US/08412376  
; Patent No. 5849900  
; GENERAL INFORMATION:  
; APPLICANT: Moelling, Karin  
; TITLE OF INVENTION: Inhibition Of Viruses By Antisense  
; TITLE OF INVENTION: Oligomers Capable Of Binding to Polypurine-Rich Tract Of Single  
; TITLE OF INVENTION: Stranded RNA Or RNA-DNA Hybrids  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz  
; ADDRESSEE: Mackiewicz & No. 5849900ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/412,376  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/954,184  
; FILING DATE: 29-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Doreen Yanko Trujillo  
; REGISTRATION NUMBER: 35,719  
; REFERENCE/DOCKET NUMBER: APOL-0021

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; ANTI-SENSE: Yes  
; US-08-412-376-31

Query Match 71.0%; Score 14.2; DB 3; Length 36;  
Best Local Similarity 84.2%; Pred. No. 28;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cttgaaggaggttcctcc 20  
|| ||||| |||||  
Db 8 CTGAAGGAGGTTCCTTC 26

RESULT 11  
US-08-434-411-1  
; Sequence 1, Application US/08434411  
; Patent No. 5681720  
; GENERAL INFORMATION:  
; APPLICANT: KUGA, TETSURO  
; APPLICANT: MIYAJI, HIROMASA  
; APPLICANT: SATO, MORIYUKI  
; APPLICANT: OKABE, MASAMI  
; APPLICANT: MORIMOTO, MAKOTO  
; APPLICANT: ITOH, SEIGA  
; APPLICANT: YAMASAKI, MOTOO  
; APPLICANT: YOKOO, YOSHIHARU  
; APPLICANT: YAMAGUCHI, KAZUO  
; APPLICANT: YOSHIDA, HAJIME  
; APPLICANT: YOSHINORI, KOMATSU  
; TITLE OF INVENTION: NOVEL POLYPEPTIDES  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/434,411  
; FILING DATE: 03-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 306799/86  
; FILING DATE: 23-DEC-1986  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 51357/88  
; FILING DATE: 04-MAR-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 80088/88  
; FILING DATE: 31-MAR-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAWFORD, ARTHUR  
; REGISTRATION NUMBER: 25327  
; REFERENCE/DOCKET NUMBER: 249-73  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)816-4000  
; TELEFAX: (703)816-4100  
; TELEX: 200797 NIXN UR

;  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 525 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..522  
; OTHER INFORMATION: /product= "HUMAN GRANULOCYTE COLONY  
; OTHER INFORMATION: STIMULATING FACTOR"  
US-08-434-411-1

Query Match 71.0%; Score 14.2; DB 1; Length 525;  
Best Local Similarity 84.2%; Pred. No. 52;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 ccttgaagggttcctc 19  
||| ||||| |||||  
Db 273 CCTGGAAGGATCTCCCC 291

RESULT 12  
US-08-434-402-1  
; Sequence 1, Application US/08434402  
; Patent No. 5714581  
; GENERAL INFORMATION:  
; APPLICANT: KUGA, TETSURO  
; APPLICANT: MIYAJI, HIROMASA  
; APPLICANT: SATO, MORIYUKI  
; APPLICANT: OKABE, MASAMI  
; APPLICANT: MORIMOTO, MAKOTO  
; APPLICANT: ITOH, SEIGA  
; APPLICANT: YAMASAKI, MOTOO  
; APPLICANT: YOKOO, YOSHIHARU  
; APPLICANT: YAMAGUCHI, KAZUO  
; APPLICANT: YOSHIDA, HAJIME  
; APPLICANT: YOSHINORI, KOMATSU  
; TITLE OF INVENTION: NOVEL POLYPEPTIDES  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/434,402  
; FILING DATE: 03-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 306799/86  
; FILING DATE: 23-DEC-1986  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 51357/88  
; FILING DATE: 04-MAR-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 80088/88  
; FILING DATE: 31-MAR-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAWFORD, ARTHUR  
; REGISTRATION NUMBER: 25327  
; REFERENCE/DOCKET NUMBER: 249-72  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)816-4000

;  
; TELEFAX: (703)816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 525 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..522  
; OTHER INFORMATION: /product= "HUMAN GRANULOCYTE COLONY  
; OTHER INFORMATION: STIMULATING FACTOR"  
US-08-434-402-1  
  
Query Match 71.0%; Score 14.2; DB 2; Length 525;  
Best Local Similarity 84.2%; Pred. No. 52;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 ccttgaagggttcctc 19  
||| ||||| |||||  
Db 273 CCTGGAAGGATCTCCCC 291  
  
RESULT 13  
US-08-783-288-1  
; Sequence 1, Application US/08783288  
; Patent No. 5795968  
; GENERAL INFORMATION:  
; APPLICANT: KUGA, TETSURO  
; APPLICANT: MIYAJI, HIROMASA  
; APPLICANT: SATO, MORIYUKI  
; APPLICANT: OKABE, MASAMI  
; APPLICANT: MORIMOTO, MAKOTO  
; APPLICANT: ITOH, SEIGA  
; APPLICANT: YAMASAKI, MOTOO  
; APPLICANT: YOKOO, YOSHIHARU  
; APPLICANT: YAMAGUCHI, KAZUO  
; APPLICANT: YOSHIDA, HAJIME  
; APPLICANT: YOSHINORI, KOMATSU  
; TITLE OF INVENTION: NOVEL POLYPEPTIDES  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/783,288  
; FILING DATE: 10-JAN-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/434,411  
; FILING DATE: 03-MAY-1995  
; APPLICATION NUMBER: JP 306799/86  
; FILING DATE: 23-DEC-1986  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 51357/88  
; FILING DATE: 04-MAR-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 80088/88  
; FILING DATE: 31-MAR-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAWFORD, ARTHUR

; REGISTRATION NUMBER: 25327  
; REFERENCE/DOCKET NUMBER: 249-73  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)816-4000  
; TELEFAX: (703)816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 525 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..522  
; OTHER INFORMATION: /product= "HUMAN GRANULOCYTE COLONY  
; OTHER INFORMATION: STIMULATING FACTOR"  
; US-08-783-288-1

Query Match 71.0%; Score 14.2; DB 2; Length 525;  
Best Local Similarity 84.2%; Pred. No. 52;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaagggttcctc 19  
||| ||||| |||||  
Db 273 CCTGGAAGGATCTCCCC 291

RESULT 14  
US-08-890-640-1  
; Sequence 1, Application US/08890640  
; Patent No. 5994518  
; GENERAL INFORMATION:  
; APPLICANT: KUGA, TETSURO  
; APPLICANT: MIYAJI, HIROMASA  
; APPLICANT: SATO, MORIYUKI  
; APPLICANT: OKABE, MASAMI  
; APPLICANT: MORIMOTO, MAKOTO  
; APPLICANT: ITOH, SEIGA  
; APPLICANT: YAMASAKI, MOTOO  
; APPLICANT: YOKOO, YOSHIHARU  
; APPLICANT: YAMAGUCHI, KAZUO  
; APPLICANT: YOSHIDA, HAJIME  
; APPLICANT: YOSHINORI, KOMATSU  
; TITLE OF INVENTION: NOVEL POLYPEPTIDES  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHVE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/890,640  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/434,411  
; FILING DATE: 03-MAY-1995  
; APPLICATION NUMBER: JP 306799/86  
; FILING DATE: 23-DEC-1986  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 51357/88  
; FILING DATE: 04-MAR-1988  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 80088/88  
; FILING DATE: 31-MAR-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAWFORD, ARTHUR  
; REGISTRATION NUMBER: 25327  
; REFERENCE/DOCKET NUMBER: 249-73  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)816-4000  
; TELEFAX: (703)816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 525 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..522  
; OTHER INFORMATION: /product= "HUMAN GRANULOCYTE COLONY  
; OTHER INFORMATION: STIMULATING FACTOR"  
; US-08-890-640-1

Query Match 71.0%; Score 14.2; DB 4; Length 525;  
Best Local Similarity 84.2%; Pred. No. 52;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaagggttcctc 19  
||| ||||| |||||  
Db 273 CCTGGAAGGATCTCCCC 291

RESULT 15  
PCT-US94-12873-1  
; Sequence 1, Application PC/TUS9412873  
; GENERAL INFORMATION:  
; APPLICANT:  
; APPLICANT:  
; APPLICANT:  
; TITLE OF INVENTION: HYBRID CYTOKINES  
; NUMBER OF SEQUENCES: 26  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" diskette, 1.44Mb, double side, high density  
; COMPUTER: AST-IBM Compatible  
; OPERATING SYSTEM: MS-DOS Version 6  
; SOFTWARE: WORD for WINDOWS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/12873  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 525  
; TYPE: nucleic acid  
; STRANDEDNESS: double stranded  
; TOPOLOGY: linear  
; PCT-US94-12873-1

Query Match 71.0%; Score 14.2; DB 6; Length 525;  
Best Local Similarity 84.2%; Pred. No. 52;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccttgaagggttcctc 19  
||| ||||| |||||  
Db 273 CCTGGAAGGATCTCCCC 291

Search completed: May 23, 2000, 09:40:21  
Job time: 8760 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:39:11 ; Search time 693.2 Seconds  
(without alignments)  
-28.067 Million cell updates/sec

Title: US-08-945-805-2  
Perfect score: 20  
Sequence: 1 ttgccgtacctgacttagcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	16.8	84.0	174795	32 HSA388K2	AL121584 Homo sapi
C 2	16.8	84.0	184675	57 AC012018	AC012018 Homo sapi
C 3	16.4	82.0	299081	41 AC006892	AC006892 Caenorhab
C 4	15.8	79.0	39	5 AR009260	AR009260 Sequence
C 5	15.8	79.0	42729	1 MTCY71	Z92771 Mycobacteri
C 6	15.8	79.0	136205	33 AL137860	AL137860 Homo sapi
C 7	15.8	79.0	158274	33 AL138914	AL138914 Homo sapi
C 8	15.8	79.0	170640	54 AC021586	AC021586 Homo sapi
C 9	15.8	79.0	197045	54 AC016821	AC016821 Homo sapi
C 10	15.4	77.0	2207	7 CPTK10	Z46647 C.plantagin
C 11	15.4	77.0	6555	2 AF155197	AF155197 Acetivibr
C 12	15.4	77.0	86408	41 AC007952	AC007952 Homo sapi
C 13	15.4	77.0	128550	33 HSAJ9617	AJ009617 Homo sapi
C 14	15.4	77.0	173219	44 AC010362	AC010362 Homo sapi
C 15	15.4	77.0	184541	42 AC012246	AC012246 Homo sapi
C 16	15.4	77.0	218807	45 AC008571	AC008571 Homo sapi
C 17	15.2	76.0	315	34 CYGMTCOIF	D30635 Calyptogena
C 18	15.2	76.0	315	34 CYGMTCOIH	D30637 Calyptogena
C 19	15.2	76.0	660	8 CNS0197S	AL111296 Botrytis
C 20	15.2	76.0	861	12 AF025506	AF025506 Rattus no
C 21	15.2	76.0	2746	10 HSA010089	A010069 Homo sapi
C 22	15.2	76.0	3014	35 PLU81824	U81824 Pacifastacu
C 23	15.2	76.0	4018	5 E06984	E06984 DNA encodin
C 24	15.2	76.0	4199	10 HS510H16B	AL079310 Novel hum
C 25	15.2	76.0	4470	2 TFENTRBC	L18975 Thiobacillu
C 26	15.2	76.0	4699	1 ECOTSH	L27423 Escherichia
C 27	15.2	76.0	5984	34 CYNNCHANN	L15445 Cyanea capi
C 28	15.2	76.0	7493	12 MUS2PF36G	L42317 Mus musculu
C 29	15.2	76.0	8359	16 REFMLVCG	X02794 Friend muri
C 30	15.2	76.0	24643	8 SPAC6C3	Z69731 S.pombe chr
C 31	15.2	76.0	41009	32 CEY94A7	Z99294 Caenorhabdi
C 32	15.2	76.0	76243	55 AC023489	AC023489 Trypanoso
C 33	15.2	76.0	83328	11 HS386C22	Z33784 Human DNA s
C 34	15.2	76.0	94757	41 AF164115	AF164115 Homo sapi
C 35	15.2	76.0	100175	33 AL137858	AL137858 Homo sapi
C 36	15.2	76.0	101714	43 AC017528	AC017528 Drosophil
C 37	15.2	76.0	102484	11 HS510H16	AL008635 Human DNA
C 38	15.2	76.0	105236	52 AC009534	AC009534 Drosophil
C 39	15.2	76.0	105320	32 AF000668	AF000668 Homo sapi
C 40	15.2	76.0	130027	40 AC004982	AC004982 Homo sapi
C 41	15.2	76.0	145295	55 AC021778	AC021778 Homo sapi
C 42	15.2	76.0	145597	16 AF030027	AF030027 Equine he
C 43	15.2	76.0	145880	11 HS302D9	Z82198 Human DNA s
C 44	15.2	76.0	159515	12 AC002327	AC002327 Mus muscu
C 45	15.2	76.0	160367	40 AC009542	AC009542 Homo sapi

ALIGNMENTS

```

RESULT 1
HSBA388K2/c
LOCUS HSBA388K2 174795 bp DNA HTG 07-JAN-2000
DEFINITION Homo sapiens chromosome 20 clone RP11-388K2, *** SEQUENCING IN
PROGRESS ***, 11 unordered pieces.
ACCESSION AL121584
VERSION AL121584.9 GI:6981980
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174795)
Direct Submission
Submitted (07-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 16, 2000 this sequence version replaced gi:6468360.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00005 Length: 22176bp
Contig_ID: 00083 Length: 15389bp
Contig_ID: 00357 Length: 2936bp
Contig_ID: 00915 Length: 1067bp
Contig_ID: 01293 Length: 21227bp
Contig_ID: 01458 Length: 1342bp
Contig_ID: 01720 Length: 11333bp
Contig_ID: 01936 Length: 34797bp
Contig_ID: 01958 Length: 13024bp
Contig_ID: 02088 Length: 19321bp
Contig_ID: 02121 Length: 24183bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 22176: contig of 22176 bp in length
* 22177 22976: gap of 800 bp
* 22977 38365: contig of 15389 bp in length
* 38366 39165: gap of 800 bp
* 39166 42101: contig of 2936 bp in length
* 42102 42901: gap of 800 bp
* 42902 43968: contig of 1067 bp in length
* 43969 44768: gap of 800 bp
* 44769 65995: contig of 21227 bp in length
* 65996 66795: gap of 800 bp
* 66796 68137: contig of 1342 bp in length
* 68138 68937: gap of 800 bp
* 68938 80270: contig of 11333 bp in length
* 80271 81070: gap of 800 bp
* 81071 115867: contig of 34797 bp in length
* 115868 116667: gap of 800 bp
* 116668 129691: contig of 13024 bp in length
* 129692 130491: gap of 800 bp
* 130492 149812: contig of 19321 bp in length
* 149813 150612: gap of 800 bp
* 150613 174795: contig of 24183 bp in length.
Location/Qualifiers
1. .174795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP11-388K2"
/clone_lib="RPC1-11.2"
FEATURES
source

```

```

BASE COUNT 50838 a 32699 c 33370 g 49887 t 8001 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 32; Length 174795;
Best Local Similarity 90.0%; Pred. No. 86;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgcctacctgacttagcc 20
||||| |||||||||
Db 58579 TTGCCTTACCTGACTTAGAC 58560

RESULT 2
AC012018/c
LOCUS AC012018 184675 bp DNA HTG 19-FEB-2000
DEFINITION Homo sapiens clone RP11-691B3, WORKING DRAFT SEQUENCE, 41 unordered
pieces.
ACCESSION AC012018
VERSION AC012018.4 GI:7007751
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184675)
Muzay, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,
Bodota, B., Bouch, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,
Dugan-Rocha, S., Durbin, K. J., Fernandez, C., Ferraguto, D.,
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J. H., Gorrell, L. L.,
Guevara, W., Harris, K., Hernandez, J., Hodgeson, A., Hogues, M.,
Holloway, C., Hosak, H., Jackson, L. E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R. J., Lu, J.,
Luchter, R., Martin, R., Martinez, C., McLeod, M. P., Mei, G., Morgan, M.,
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L. L.,
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R.,
Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M.,
Watlington, S., Weinstein, G., Weinstein, I. R., Williamson, A.,
Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and
Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 184675)
Worley, K.C.
Direct Submission
Submitted (19-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 19, 2000 this sequence version replaced gi:6692184.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMOB
Center clone name: RP11-691B3
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 150223 bases at least Q40
Consensus quality: 162960 bases at least Q30
Consensus quality: 169717 bases at least Q20
Estimated insert size: 184450; agarose-fp estimation
Estimated insert size: 170396; sum-of-contigs estimation
Quality coverage: 3.1x in Q20 bases; agarose-fp estimation
Quality coverage: 3.3x in Q20 bases; sum-of-contigs estimation

```

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 41 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 798: contig of 798 bp in length  
\* 799 818: gap of unknown length  
\* 819 1813: contig of 995 bp in length  
\* 1814 1833: gap of unknown length  
\* 1834 2617: contig of 784 bp in length  
\* 2618 2637: gap of unknown length  
\* 2638 3431: contig of 794 bp in length  
\* 3432 3451: gap of unknown length  
\* 3452 4294: contig of 843 bp in length  
\* 4295 4314: gap of unknown length  
\* 4315 5079: contig of 765 bp in length  
\* 5080 5099: gap of unknown length  
\* 5100 6213: contig of 1114 bp in length  
\* 6214 6233: gap of unknown length  
\* 6234 6572: contig of 339 bp in length  
\* 6573 6592: gap of unknown length  
\* 6593 7382: contig of 790 bp in length  
\* 7383 7402: gap of unknown length  
\* 7403 7570: contig of 168 bp in length  
\* 7571 7590: gap of unknown length  
\* 7591 8829: contig of 1239 bp in length  
\* 8830 8849: gap of unknown length  
\* 8850 10320: contig of 1471 bp in length  
\* 10321 10340: gap of unknown length  
\* 10341 11132: contig of 792 bp in length  
\* 11133 11152: gap of unknown length  
\* 11153 12068: contig of 916 bp in length  
\* 12069 12088: gap of unknown length  
\* 12089 13047: contig of 959 bp in length  
\* 13048 13067: gap of unknown length  
\* 13068 13862: contig of 795 bp in length  
\* 13863 13882: gap of unknown length  
\* 13883 14741: contig of 859 bp in length  
\* 14742 14761: gap of unknown length  
\* 14762 15704: contig of 943 bp in length  
\* 15705 15724: gap of unknown length  
\* 15725 17204: contig of 1480 bp in length  
\* 17205 17224: gap of unknown length  
\* 17225 18720: contig of 1496 bp in length  
\* 18721 18740: gap of unknown length  
\* 18741 20395: contig of 1655 bp in length  
\* 20396 20415: gap of unknown length  
\* 20416 22333: contig of 1918 bp in length  
\* 22334 22353: gap of unknown length  
\* 22354 23812: contig of 1459 bp in length  
\* 23813 23832: gap of unknown length  
\* 23833 25585: contig of 1753 bp in length  
\* 25586 25605: gap of unknown length  
\* 25606 27411: contig of 1806 bp in length  
\* 27412 27431: gap of unknown length  
\* 27432 28350: contig of 919 bp in length  
\* 28351 28370: gap of unknown length  
\* 28371 30149: contig of 1779 bp in length  
\* 30150 30169: gap of unknown length  
\* 30170 32516: contig of 2347 bp in length  
\* 32517 32536: gap of unknown length  
\* 32537 33532: contig of 996 bp in length  
\* 33533 33552: gap of unknown length  
\* 33553 36955: contig of 3403 bp in length  
\* 36956 36975: gap of unknown length  
\* 36976 40842: contig of 3867 bp in length  
\* 40843 40862: gap of unknown length  
\* 40863 44819: contig of 3957 bp in length  
\* 44820 44839: gap of unknown length

\* 44840 49429: contig of 4590 bp in length  
\* 49430 49449: gap of unknown length  
\* 49450 56085: contig of 6636 bp in length  
\* 56086 56105: gap of unknown length  
\* 56106 65998: contig of 9893 bp in length  
\* 65999 66018: gap of unknown length  
\* 66019 76555: contig of 10537 bp in length  
\* 76556 76575: gap of unknown length  
\* 76576 88270: contig of 11695 bp in length  
\* 88271 88290: gap of unknown length  
\* 88291 101042: contig of 12752 bp in length  
\* 101043 101062: gap of unknown length  
\* 101063 118515: contig of 17453 bp in length  
\* 118516 118535: gap of unknown length  
\* 118536 132271: contig of 13736 bp in length  
\* 132272 132291: gap of unknown length  
\* 132292 184675: contig of 52384 bp in length.

#### FEATURES

Location/Qualifiers  
1..184675  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-691B3"

BASE COUNT 52548 a 37526 c 36626 g 56628 t 1347 others  
ORIGIN

Query Match 84.0%; Score 16.8; DB 57; Length 184675;  
Best Local Similarity 90.0%; Pred. No. 86;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgcccgtacctgacttagcc 20  
||||| | ||||| |||||

Db 25777 TTGCAGGACCTGACTTAGCC 25758

#### RESULT 3

AC006892

LOCUS

DEFINITION

AC006892

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Mar 1, 1999 this sequence version replaced gi:4263456.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 10 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 3087: contig of 3087 bp in length

\* 3088 3101: gap of unknown length

\* 3102 7168: contig of 4067 bp in length

\* 7169 7182: gap of unknown length

\* 7183 16968: contig of 9786 bp in length

\* 16969 16982: gap of unknown length

\* 16983 30313: contig of 13331 bp in length

AC006892 299081 bp DNA HTG 26-FEB-1999  
Caenorhabditis elegans clone Y69A2, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
10 unordered pieces.

AC006892

AC006892.2 GI:4309907

HTG; HTGS-PHASE1.

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

Waterston, R.H.

The sequence of Caenorhabditis elegans clone

2 (bases 1 to 299081)

Waterston, R.H.

Direct Submission

Submitted (24-FEB-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA



```
/gene="Rv3262"  
/note="Rv3262, (MTCY71.02, unknown), len: 448 aa, similar  
to AE000874_11 Methanobacterium thermoautotro 027098  
CONSERVED PROTEIN (252 aa; E(): 7.6e-18; 37.3% identity in  
249 aa overlap). Also similar to Q58178 hypothetical 27.1  
kDa protein 0768 (249 aa) opt: 375, E(): 6.9e-17, (33.9%  
identity in 239 aa overlap)"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein Rv3262"  
/protein_id="CAB07089.1"  
/db_xref="GI:1877317"  
/db_xref="SPTREMBL:P96867"  
/translation="MTPEHGSASTIEILVIGLPEPRPGDLSAAVAAPWLRDGD  
VYVTSVSKVCEGLVPAPEDQDRLRKLTIEDEAVRLVARKDRTLTENRGLIV  
QAAAGVSGNKGSEIALLPVDPDASATLRAGLRRLRVTVAVIITDMGRAWNRQ  
DTAAGVAGNAGVLRNYPGVNDELVTTEVAADETIAAADLVKGLTATPVAVVR  
GFGVSDGSTARQLLRPGANDLFWLGTAEALRQQAQLLRSVRFPSTDPVPGDLY  
EAAVAELTAPAPHTTRPTFVWLQTPAIRLRDLDRKWSDLTSDGLPADAIEER  
VARGQILDAPEVPIVPLVPGAHSYPDAAITDAETMTFTVAAGAAQALLVALVRG  
LGSCWIGSTIFAADLVDELDPVDWEPLGAITAIGYADEPSGLRDPVPAADLLIK"  
2856. 4517  
/gene="Rv3263"  
/note="Rv3263, (MTCY71.03), len: 553, DNA methylase,  
(alternative start at aa 25), similar to eg MT57_ECOLI  
p25240 modification methylase eco57i (544 aa), fasta  
scores, opt: 595, E(): 0, (30.4% identity in 507 aa  
overlap); contains P50092 N-6 Adenine-specific DNA  
methylases signature"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein Rv3263"  
/protein_id="CAB07090.1"  
/db_xref="GI:1877318"  
/db_xref="SPTREMBL:P96868"  
/translation="MOPSHPTREGAVIRYVGSSLDTCPTTFAGTKAASADKVRGGY  
TPPAVFLAHVHWAGPKILEPCGDGRILRELSTADHAHGVELVAREAKSRDFA  
SVDPENLFTWLKTLGSDWGAGNPPYIRFGNWSQRDPALFMRVGLRPTKLTN  
AWPFFVASTLRDGRGLVLPVPAELLQTYAQLREFLLRSYREITLTFERLVD  
GILQVFLFCGVGPGPAHRTVRGLGANDLNLGDKDTNESAPALLHEKERTKYP  
LDPAQIRLLGLSATMIRGLADVDVIGVTRGNSFFFTDAKAQALGLRAHCVP  
VPSAQLSLGTYSDRCRAGVNHRLDADYPTDPAIVHITAGEAAGVHLGYK  
CSIRKPWSPSLMPOLEMLRQIHFAPLRTVNAAAATSTDTVHRVLDPNVDPATLA  
AVFNHSAFAEATMGSRYSGGILELEPREAQOLPMPPAVGSAAELAQDVLKANE  
IDKALDVEDRVLIDGLSPRLVAGCRAAWLTDRRTKRSRR"  
3225. 3245  
/gene="Rv3263"  
/note="P50092 N-6 Adenine-specific DNA methylases  
signature"  
complement(4577. .5656)  
/gene="rmlA2"  
complement(4577. .5656)  
/note="Rv3264C, (MTCY71.04C), len: 359, rfbF transferase,  
similar to MPG1_YEAST P41940 mannose-1-phosphate  
guanylyltransferase (361 aa), fasta scores, opt: 557, E():  
1.9e-27, (28.9% identity in 353 aa overlap), also similar  
to RBAL_ECOLIP37744, glucose-1-phosphate  
thymidyllyltransferase (293 aa), fasta scores, opt: 317,  
E(): 7.7e-13, (26.4% identity in 235 aa overlap)"  
/codon_start=1  
/transl_table=11  
/product="rmlA2"  
/protein_id="CAB07091.1"  
/db_xref="GI:1877319"  
/db_xref="SPTREMBL:P96869"  
/translation="MATHQVDVAVLVGSGKTRLRPLTSLAPKMLTAGLPFLTHLS  
RIAAAGIEHVLIGTSYKPAVFAEFGSGALGLQIEVTHEPLGTCGGTIANVAGLR  
NDTAMVNGDVLGSAQLQDLDFHRSNRADTLQLVRGDPGPAFCGGTDEEDRVAF  
LEKTDPTDQINAGCVFERNVIDRIPOQREVREVFALLADGDKITGYVDAS  
YWRDMGTPEDFVRGSADLVIRGIAPSPALRGHRGQLVHDGAASPGLALLGGTVVGRG
```

```
AEIGPGRFLGAVIFDQVRVEAGCVIERSIIGFARIGPRALIRDGVIGDADIGARC  
ELLSGARWPCVFLPDGGIRYSDV"  
complement(5658. .6563)  
/gene="wbbL"  
complement(5658. .6563)  
/note="wbbL" (MTCY71.05C), len: 301, unknown but  
similar to YEFJ_ECOLI P36667 hypothetical 31.1 kd protein  
in gnd-r (264 aa), fasta scores, opt: 253, E(): 9.8e-10,  
(28.6% identity in 259 aa overlap)"  
/codon_start=1  
/transl_table=11  
/product="wbbL"  
/protein_id="CAB07092.1"  
/db_xref="GI:1877320"  
/db_xref="SPTREMBL:P96870"  
/translation="MVAVTYSPGHLERFLASLATERPVSVLLADNGSDGTPOAA  
VQRYPNVRLPTGANLGYTAVNRTIAQLGEMAGDAGEPWVDVIVANPVQVQPGS  
IDALLDAASRWPRAGALGPLIPDGSVYPSARQMPSLIRGMHVLGPFVPRNPWT  
AYROERLEPSERPYGLSGCLLVRSFAGQGVGFDERYFMYMEDVDLGRKAGWL  
SVYVPSAEVLHKAHSTGRDPASHLAAHKKSTYIFLADRHSGWNRAPLRMTLRGSLAL  
RSHLMVRSSLLRRSRRLKLVGRH"  
complement(6574. .7488)  
/gene="rmlD"  
complement(6574. .7488)  
/note="rmlD" (MTCY71.06C), len: 304, rfbD, similar to  
eg STRL_STRGR P29781 dtdp-4-dehydrorhamnose reductase (304  
aa), fasta scores, opt: 788, E(): 0, (47.4% identity in  
304 aa overlap)"  
/codon_start=1  
/transl_table=11  
/product="rmlD"  
/protein_id="CAB07093.1"  
/db_xref="GI:1877321"  
/db_xref="SPTREMBL:P96871"  
/translation="MAGRSERLVTGAGQGLGSHLTAQAAREGRDMLALTSSQWDITD  
PAAERTIRHGDDVINCAVTDVDAESNEAVAVNATVQPHLARACARVARGLIHV  
STDYVFDGFGGAEPRPYETDPAQGVYARSLAGAEQVLAAPFAAVVTVAVVT  
GGTCKDFAVMRRLAAGHRVDVDDQTSPTYVADLAEALLADAGVGRVLRHAAN  
EGVYSRFGQARAAPFECEGADQVRVPSVSSAQFPAPRASSYSALSSQWALAGLTPLR  
HWSALATAAAPANSTSIDRLSPSTRD"  
7564. .9060  
/gene="Rv3267"  
Query Match 79.0%; Score 15.8; DB 1; Length 42729;  
Best Local Similarity 89.5%; Pred. No. 3.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 2 tgcqgtacctgacttagcc 20  
|||||l|||||l  
Db 22484 TGCCGTACCCGACTTGGCC 22466  
RESULT 6  
AL137860/c  
LOCUS AL137860 136205 bp DNA HTG 30-JAN-2000  
DEFINITION Homo sapiens chromosome 1 clone RP4-633N17 map p34.1-34.3, ***  
SEQUENCING IN PROGRESS ***, 29 unordered pieces.  
ACCESSION AL137860  
VERSION AL137860.1 GI:6982209  
KEYWORDS HTG; HTGS_PHASE1.  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 136205)  
AUTHORS Pavitt, R.  
TITLE Direct Submission  
JOURNAL Submitted (30-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
COMMENT On Feb 16, 2000 this sequence version replaced gi:6840766.
```

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Contig\_ID: 00007 Length: 3929bp

- Contig\_ID: 00010 Length: 7615bp
- Contig\_ID: 00037 Length: 2258bp
- Contig\_ID: 00077 Length: 1127bp
- Contig\_ID: 00160 Length: 1040bp
- Contig\_ID: 00176 Length: 3052bp
- Contig\_ID: 00283 Length: 1602bp
- Contig\_ID: 00460 Length: 3112bp
- Contig\_ID: 00478 Length: 1870bp
- Contig\_ID: 00488 Length: 1873bp
- Contig\_ID: 00490 Length: 9326bp
- Contig\_ID: 00502 Length: 2089bp
- Contig\_ID: 00516 Length: 986bp
- Contig\_ID: 00554 Length: 3846bp
- Contig\_ID: 00572 Length: 1582bp
- Contig\_ID: 00605 Length: 2564bp
- Contig\_ID: 00607 Length: 1336bp
- Contig\_ID: 00668 Length: 2215bp
- Contig\_ID: 00726 Length: 17464bp
- Contig\_ID: 00733 Length: 3719bp
- Contig\_ID: 00762 Length: 4206bp
- Contig\_ID: 00846 Length: 1542bp
- Contig\_ID: 00851 Length: 1588bp
- Contig\_ID: 00904 Length: 6562bp
- Contig\_ID: 00975 Length: 3252bp
- Contig\_ID: 01007 Length: 5929bp
- Contig\_ID: 01016 Length: 3961bp
- Contig\_ID: 01024 Length: 6520bp
- Contig\_ID: 01083 Length: 7640bp.

\* NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

- 1 3929: contig of 3929 bp in length
- 3930 4729: gap of 800 bp
- 4730 12344: contig of 7615 bp in length
- 12345 13144: gap of 800 bp
- 13145 15402: contig of 2258 bp in length
- 15403 16202: gap of 800 bp
- 16203 17329: contig of 1127 bp in length
- 17330 18129: gap of 800 bp
- 18130 19169: contig of 1040 bp in length
- 19170 19969: gap of 800 bp
- 19970 23021: contig of 3052 bp in length
- 23022 23821: gap of 800 bp
- 23822 25423: contig of 1602 bp in length
- 25424 26223: gap of 800 bp
- 26224 29335: contig of 3112 bp in length
- 29336 30135: gap of 800 bp
- 30136 32005: contig of 1870 bp in length
- 32006 32805: gap of 800 bp
- 32806 34678: contig of 1873 bp in length
- 34679 35478: gap of 800 bp
- 35479 44804: contig of 9326 bp in length
- 44805 45604: gap of 800 bp
- 45605 47693: contig of 2089 bp in length
- 47694 48493: gap of 800 bp
- 48494 49479: contig of 986 bp in length
- 49480 50279: gap of 800 bp
- 50280 54125: contig of 3846 bp in length
- 54126 54925: gap of 800 bp
- 54926 56507: contig of 1582 bp in length
- 56508 57307: gap of 800 bp

- \* 57308 59871: contig of 2564 bp in length
- \* 59872 60671: gap of 800 bp
- \* 60672 62007: contig of 1336 bp in length
- \* 62008 62807: gap of 800 bp
- \* 62808 65022: contig of 2215 bp in length
- \* 65023 65022: gap of 800 bp
- \* 65823 83286: contig of 17464 bp in length
- \* 83287 84086: gap of 800 bp
- \* 84087 87805: contig of 3719 bp in length
- \* 87806 88605: gap of 800 bp
- \* 88606 92811: contig of 4206 bp in length
- \* 92812 93611: gap of 800 bp
- \* 93612 95153: contig of 1542 bp in length
- \* 95154 95953: gap of 800 bp
- \* 95954 97541: contig of 1588 bp in length
- \* 97542 98341: gap of 800 bp
- \* 98342 104903: contig of 6562 bp in length
- \* 104904 105703: gap of 800 bp
- \* 105704 108955: contig of 3252 bp in length
- \* 108956 109755: gap of 800 bp
- \* 109756 115684: contig of 5929 bp in length
- \* 115685 116484: gap of 800 bp
- \* 116485 120445: contig of 3961 bp in length
- \* 120446 121245: gap of 800 bp
- \* 121246 127766: contig of 6520 bp in length
- \* 127766 128565: gap of 800 bp
- \* 128566 136205: contig of 7640 bp in length.

FEATURES  
source

1. .136205  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p34.1-34.3"  
/clone="RP4-633N17"  
/clone\_lib="RPCI-4"

BASE COUNT 27933 a 28844 c 28528 g 28495 t 22405 others  
ORIGIN

Query Match 79.0%; Score 15.8; DB 33; Length 136205;  
Best Local Similarity 89.5%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcgcgtacctgacttagcc 20  
|||||

Db 133251 TGCCTTACTGTACTTGGC 133233

RESULT 7

AL138914/c

LOCUS AL138914 158274 bp DNA HTG 08-FEB-2000  
DEFINITION Homo sapiens chromosome 13 clone RP11-15L1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 7 unordered pieces.

ACCESSION AL138914

VERSION AL138914.2 GI:6982929

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 158274)

AUTHORS Burton,J.

TITLE Direct Submission

JOURNAL Submitted (08-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

COMMENT On Feb 16, 2000 this sequence version replaced gi:6966888

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate

segments. Contig\_ID: 00079 Length: 23467bp  
Contig\_ID: 00580 Length: 15043bp  
Contig\_ID: 00713 Length: 24867bp  
Contig\_ID: 01203 Length: 19647bp  
Contig\_ID: 01303 Length: 33285bp  
Contig\_ID: 01305 Length: 7968bp  
Contig\_ID: 01402 Length: 29197bp.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 23467: contig of 23467 bp in length  
\* 23468 24267: gap of 800 bp  
\* 24268 39310: contig of 15043 bp in length  
\* 39311 40110: gap of 800 bp  
\* 40111 64977: contig of 24867 bp in length  
\* 64978 65777: gap of 800 bp  
\* 65778 85424: contig of 19647 bp in length  
\* 85425 86224: gap of 800 bp  
\* 86225 119509: contig of 33285 bp in length  
\* 119510 120309: gap of 800 bp  
\* 120310 128277: contig of 7968 bp in length  
\* 128278 129077: gap of 800 bp  
\* 129078 158274: contig of 29197 bp in length.

FEATURES

source  
1. 158274  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-15L1"  
/clone\_lib="RPC1-11.1"  
BASE COUNT 48106 a 29626 c 28743 g 46985 t 4814 others  
ORIGIN

Query Match 79.0%; Score 15.8; DB 33; Length 158274;  
Best Local Similarity 89.5%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ttgcccgtacctgacttagc 19

Db 113988 TTGCCATCCCTGACTTACG 113970

RESULT 8

AC021586/c

LOCUS

DEFINITION Homo sapiens clone RP11-112B10, WORKING DRAFT SEQUENCE, 27

unordered pieces.

ACCESSION AC021586

VERSION AC021586.3 GI:6939641

KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 170640)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Unpublished

2 (bases 1 to 170640)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bada, F.,

Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,

Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

Dearellano, K., Dewar, K., Domino, N., Doyle, M., Fenestor, J.,

Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,

Gardyna, S., Grant, G., Hagsos, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrim, J., Menus, L., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 8, 2000 this sequence version replaced gi:6899710.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5768

Center clone name: 112\_B\_10

----- Summary Statistics

Sequencing vector: M13; M77815: 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158722 bases at least Q40

Consensus quality: 166090 bases at least Q30

Consensus quality: 168861 bases at least Q20

Insert size: 184000; agarose-fp

Insert size: 170640; sum-of-contigs

Quality coverage: 4.1 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently

\* consists of 27 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1361: contig of 1361 bp in length

\* 1362 gap of unknown length

\* 2587: contig of 1226 bp in length

\* 2588 gap of unknown length

\* 3876: contig of 1289 bp in length

\* 3877 gap of unknown length

\* 5460: contig of 1584 bp in length

\* 5461 gap of unknown length

\* 6908: contig of 1448 bp in length

\* 6909 gap of unknown length

\* 8532: contig of 1624 bp in length

\* 8533 gap of unknown length

\* 10374: contig of 1842 bp in length

\* 10375 gap of unknown length

\* 12781: contig of 2407 bp in length

\* 12782 gap of unknown length

\* 15253: contig of 2472 bp in length

\* 15254 gap of unknown length

\* 18299: contig of 3046 bp in length

\* 18300 gap of unknown length

\* 20400: contig of 2101 bp in length

\* 20401 gap of unknown length

\* 23450: contig of 3050 bp in length

\* 23451 gap of unknown length

\* 27597: contig of 4147 bp in length

\* 27598 gap of unknown length

\* 30754: contig of 3157 bp in length

\* 30755 gap of unknown length

\* 35086: contig of 4332 bp in length

```
* * 35087 40111: contig of 5025 bp in length
* * gap of unknown length
* * 40112 44385: contig of 4274 bp in length
* * gap of unknown length
* * 44386 50132: contig of 5747 bp in length
* * gap of unknown length
* * 50133 56396: contig of 6264 bp in length
* * gap of unknown length
* * 56397 64410: contig of 8014 bp in length
* * gap of unknown length
* * 64411 73668: contig of 9258 bp in length
* * gap of unknown length
* * 73669 84766: contig of 11098 bp in length
* * gap of unknown length
* * 84767 93359: contig of 8593 bp in length
* * gap of unknown length
* * 93360 105975: contig of 12616 bp in length
* * gap of unknown length
* * 105976 122819: contig of 16844 bp in length
* * gap of unknown length
* * 122820 143349: contig of 20530 bp in length
* * gap of unknown length
* * 143350 170640: contig of 27291 bp in length.
* * Location/Qualifiers
* * 1. .170640
* * /organism="Homo sapiens"
* * /db_xref="taxon:9606"
* * /clone="RP11-112B10"
* * /clone_lib="RPCI-11 Human Male BAC"
* * 46695 a 39063 c 37472 g 47410 t
* *
* * Query Match 79.0%; Score 15.8; DB 54; Length 170640;
* * Best Local Similarity 89.5%; Pred. No. 3e+02;
* * Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
* *
* * QY 1 ttgccgtacctgacttagc 19
* * |||| ||||||| |||||
* * Db 107475 TTGCTGTACTGACTTAGC 107457
* *
* * RESULT 9
* * AC016821/c
* * LOCUS Homo sapiens clone RP11-404C6, WORKING DRAFT SEQUENCE, 18 unordered
* * DEFINITION pieces.
* * ACCESSION AC016821
* * VERSION AC016821.3 GI:6910507
* * KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
* * SOURCE human.
* * ORGANISM Homo sapiens
* * Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
* * Eutheria; Primates; Catarrhini; Hominidae; Homo.
* * 1 (bases 1 to 197045)
* * Smith,D.R.
* * Title Genome Therapeutics Corporation Sequencing Center: Human Genome
* * Sequence Data
* * Unpublished
* * 2 (bases 1 to 197045)
* * Smith,D.R.
* * Title Direct Submission
* * Submitted (08-DEC-1999) Genome Therapeutics Corporation, 100 Beaver
* * Street, Waltham, MA 02453, USA
* * COMMENT On Feb 6, 2000 this sequence version replaced gi:6682175.
* *
* * ----- Genome Center
* * Center: Genome Therapeutics Corporation
* * Center code: GTC
* * Web site: http://www.genomecorp.com/
* * Contact: gtc-seqcenter@genomecorp.com
* * ----- Project Information
* *
* *
* * Center project name: hg066
* * ----- Summary Statistics
* * Sequencing vector: N/A
* * Chemistry: Dye-terminator Big Dye; 100% of reads
* * Assembly program: Phrap; version 990315
* * Consensus quality: 173089 bases at least Q40
* * Consensus quality: 187039 bases at least Q30
* * Consensus quality: 189399 bases at least Q20
* * Insert size: 197045; sum-of-contigs
* * Quality coverage: 5.0x in Q20 bases; sum-of-contigs
* * -----
* * NOTE: This is a 'working draft' sequence. It currently
* * consists of 18 contigs. The true order of the pieces
* * is not known and their order in this sequence record is
* * arbitrary. Gaps between the contigs are represented as
* * runs of N, but the exact sizes of the gaps are unknown.
* * This record will be updated with the finished sequence
* * as soon as it is available and the accession number will
* * be preserved.
* * 1 1116: contig of 1116 bp in length
* * 1117 2496: gap of unknown length
* * 2497 3542: contig of 1046 bp in length
* * 3543 4566: contig of 1024 bp in length
* * 4567 6591: contig of 2025 bp in length
* * 6592 9645: contig of 3054 bp in length
* * 9646 15025: contig of 5380 bp in length
* * 15026 22690: contig of 7665 bp in length
* * 22691 33095: contig of 10405 bp in length
* * 33096 42372: contig of 9277 bp in length
* * 42373 51229: contig of 8857 bp in length
* * 51230 63202: contig of 11973 bp in length
* * 63203 74024: contig of 10822 bp in length
* * 74025 87013: contig of 12989 bp in length
* * 87014 103235: contig of 16222 bp in length
* * 103236 122156: contig of 18921 bp in length
* * 122157 155324: contig of 33168 bp in length
* * 155325 197045: contig of 41721 bp in length.
* * Location/Qualifiers
* * 1. .197045
* * /organism="Homo sapiens"
* * /db_xref="taxon:9606"
* * /clone="RP11-404C6"
* * /clone_lib="RPCI-11"
* * 48542 a 48410 c 47607 g 52455 t 31 others
* *
* * BASE COUNT 48542 a 48410 c 47607 g 52455 t 31 others
* * ORIGIN
* *
* * Query Match 79.0%; Score 15.8; DB 54; Length 197045;
* * Best Local Similarity 89.5%; Pred. No. 3e+02;
* * Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
* *
* * QY 1 ttgccgtacctgacttagc 19
* * ||||| || ||||| |||||
* * Db 77028 TTGCAGTTCCTGACTTAGC 77010
```

```
RESULT 10
LOCUS CPTKT10 2207 bp mRNA PLN 09-FEB-1996
DEFINITION C.plantagineum tkt10 gene for transketolase.
ACCESSION Z46647
VERSION Z46647.1 GI:664898
KEYWORDS tkt10 gene; transketolase.
ORGANISM Craterostigma plantagineum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asteridae; Gentiananae; Lamiales; Scrophulariaceae; Craterostigma.
1 (bases 1 to 2207)
Bartels,D.
Direct Submission
Submitted (10-NOV-1994) Bartels D., Max-Planck-Institut fuer
Zuchtforschung, Pflanzenzuechtung und Ertragsphysiologie,
Carl-von-Linne-Weg 10, Cologne, NRW, Germany, 50829
REFERENCE 2 (bases 1 to 2207)
AUTHORS Bernacchia,G., Schwall,G., Lottspeich,F., Salamini,F. and
Bartels,D.
TITLE The transketolase gene family of the resurrection plant
Craterostigma plantagineum: differential expression during the
rehydration phase
JOURNAL EMBO J. 14 (3), 610-618 (1995)
MEDLINE 95163594
FEATURES
source
Location/Qualifiers
1..2207
/organism="Craterostigma plantagineum"
/db_xref="taxon:4153"
/clone="Cpl-tkt10"
3..2042
/gene="Cpl-tkt10"
3..2042
/gene="Cpl-tkt10"
/EC_number="2.2.1.1"
/citation="[2]"
/codon_start=1
/product="transketolase"
/protein_id="CAA86608.1"
/db_xref="GI:664898"
/db_xref="SWISS-PROT:Q42675"
/translation="MAKTPSPSAAAAELVVKSVNTIRFLAIDAVENVKSGHPGMP
MGCAPMGHLVDFEPMKPNKPNWFRDRFVLSAGHMLLYGLHLAGYSVAVEDL
KGLQWGSKTPAHPENETPGVETGTPLOGGVGSAGVLAELAEHLGARYNKPFEMV
DHYTMILGDCQMEGISNEASSLAHNLGLKIALYDDNHITDGDTLAFTEDVGK
RFEALGHLVTVANGNDYDEIRAIKVAKSVTDKPTLIRKVAITIGFGSPNKANTYGV
HGNAIGPKAEATRONLGPVYETHFVDDVKKHSRHSISEGAELSAWNAKFAFEYKK
YKRAEALKLSIITGELPGWEKALPTYPESPGRPTRLTSHQNLNAAVALPGLIGGS
ADLTASNMAFLKSGDFQKPTPTGNLKFAGREHGMGAICNGVALHSPGLVPFSATYF
VFTDYMRAAIFRIALSKARVYVINTHDSIGIGEDGPTHPVHEHLASFAMPNILLVRP
ADGNETAGYKAVENAGRPSILSLRQKLPLQPGTSVEGVGRGGYVLSNDSKGEKP
EVIILMGTSLELAAARAGEELRKGKVRVSVLSWELFGSQSEYKEMVLVLPSEVTAR
VSVBAGSTFGWERVFGVLKGRAVGDIFRGASASAEELRYKEFGITVEAVVAAAKELC"
BASE COUNT 520 a 582 c 682 g 423 t
ORIGIN

Query Match 77.0%; Score 15.4; DB 7; Length 2207;
Best Local Similarity 94.1%; Pred. No. 6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 ccgtacctgactagcc 20
||||| |||||||||
Db 756 CCGTAAGTACTAGCC 740

RESULT 11
LOCUS AF155197/c
DEFINITION Acetivibrio cellulolyticus cellulosomal scaffoldin precursor (c1pv)
gene, complete cds.
ACCESSION AF155197
```

```
VERSION AF155197.1 GI:6249560
KEYWORDS Acetivibrio cellulolyticus.
SOURCE Acetivibrio cellulolyticus
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Acetivibrio.
REFERENCE 1 (bases 1 to 6555)
AUTHORS Ding,S.Y., Bayer,E.A., Steiner,D., Shoham,Y. and Lamed,R.
TITLE A novel cellulosomal scaffoldin from acetivibrio cellulolyticus
that contains a family 9 glycosyl hydrolase
JOURNAL J. Bacteriol. 181 (21), 6720-6729 (1999)
MEDLINE 20011339
REFERENCE 2 (bases 1 to 6555)
AUTHORS Ding,S.Y., Bayer,E.A., Steiner,D., Shoham,Y. and Lamed,R.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-1999) Department of Biological Chemistry, The
Weizmann Institute of Science, Rehovot 76100, Israel
FEATURES
source
Location/Qualifiers
1..6555
/organism="Acetivibrio cellulolyticus"
/db_xref="taxon:35830"
772..858
/gene="c1pv"
772..6519
/gene="c1pv"
/function="family-9 glycoside hydrolase"
/function="cellulose binding: family-III CBD"
/function="subunit binding: type-I cohesin-dockerin
interaction"
/function="cell anchoring: type-II cohesin-dockerin
interaction"
/codon_start=1
/transl_table=11
/product="cellulosomal scaffoldin precursor"
/protein_id="AAF06064.1"
/db_xref="GI:6249561"
/translation="MKKVISIMVAIAMLVTMFSTVNLGTASAAATQDYDKALDYAIOF
FDANKCGPDAAADNVSWRGACHTSDKGPSEDLTGGYHDAGDHVFGFLGPQAYSAPML
GWALYEYGNVMDSTGNKTKILSTLKYFSDYLLKCHPSANIFYHVQEGGPHSYWGAP
EKQTGRSTRKVDNSTAGTDVAGNSAALAIWLNKIDATYANKCLAASKTLIALA
KSKIGACYNESFEYKSYCYDDLAWASIMLYTAGGSTDTLLADAEKYVMGDKLNGSP
LKHWYTMCDVVYIPAFIKLAQMTGKATYKDALAYSFNWKSTYVYTGSTGTPKSNWG
HRAANPDSSNAKTYTLGALVGGPSNDSYVDVNOYQYSEVADYNSEFVALASY
LGTWMTSTTTKTATAPTKTPTSTGTGTFVNVDSVNGVGEQIIVPVSANVPSNG
VSTADMTITYDSKLEYVSGAAGSVINPTNFGINKADGKLAVLFDITYMTSTGIIS
TVGFANVTFKVLNSAPTGVGITGFGDKNLGNISATINAGSINGSTISPTTKTAV
TPTKTTTTPVPTKTTSTGTGTVNVDSVNGVGEQIIVPVSFANVPSNGSITADMTI
TYDSKLEYVSGDAGSVINPTNFGINKETDGKLVFLDYTNSTGYISTNGVFAKV
TFKVLNAGSGVIGITGATFGDKNLGSVATINAGSINGSIPTTKTAVPTKTTGV
TPTKTPVLIGTGTVSDSVNGVGEQIIVPVSFANIPANGISITADMTITYDSKLEYV
SGVPGSVINPDVNFNGINKETDGKLVFLDYTNSTGYISTGSGVTFKTVFKVSSGGG
TVGITGATFGDKNLGNVSATINAGSINGSIPTTKTTPVPTKTTVPTPGKTAVSS
NLKVEFFNAGTQAQNSIYKPRLTNTGSNAINLADVKLHYFTVVDGKAQTFWCDWS
PVGSSNVTFTVKMNPFTTGADQLEIAFSSAAGTAAANTSTEVGREAQKSDWTNNO
ADDSFNSSTATTYTSWDKVTAYSAGELWIGTEPNCTVPTTPTVPTTATATPTKTA
TATPTKSTPNPNAMAVAGVQGVGTVPVPTMTKVPVTVGSTADFTYDATKL
EYVSGAGSVINPDVNFNGINKADGKLVFLDYTNMATEYISKDGFANLTFPKKST
AAAGTTAAVGIAGTAGTAFGDSALKPTAVITDGKVEIIDGRAMKVVIANVSNAGSEV
VPVSTGSEVANGVSAADFTITYDATKLDYVSGAAGSVINPDVNFNGINKADGKLV
FLDVTMATEYISADGIFANLTFKTKSTAVNGDVAASIKSGSTATFGDKNLGKLVISIKD
GSVTVGNVPTKTTATATPTVPTKTTSTATATPTGTGTGTFNLSIDIVEGNPGSSVVVPV
KLSGISKNGISTADFTYDATKLEYISGDAGSVINPDVNFNGINKESDGLKLVFLD
YTMSTGYISTDGVFANLNFNIKSSAIGSKAEVSGTPTFGDSTLTPVAVKVNAGV
NVVKDDPNAPKVSIDITVKAATGTQVVVPVSVFNVNVPATGISTTDMTITYDTKLQYVS
GDAGSVINPDVNFNGINKADGKLVFLDYTNMATEYISADGKLVFLDYTNMATEYISADGKLV
AVNAEDATFGDSLSLSPVTVSVNGVNGIYVSEGTTSVGINPDFTVTTSTTAPIVKAG
FTVEIVGTTKSAVTDNSGYFEIKDVAAGTATVTKFKNALTRETIANVSTADKELSTS
ASPIILWAGDMAIGTQDGAINLEIDLEICKAFNTSSTDKAYQVGLDLNRDGAISLED
VNIIVAKHFNKVSDDY"
/gene="c1pv"
859..6516
mat_peptide
```

```
/gene="cipv"
/product="cellulosomal scaffoldin"

BASE COUNT      2340 a 1101 c 1259 g 1855 t
ORIGIN

Query Match      77.0%; Score 15.4; DB 2; Length 6555;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ttgcctgactgactta 17
|||||
Db 5032 TTGCTGTACCTGACTTA 5016

RESULT 12
AC007952
LOCUS
DEFINITION Homo sapiens chromosome 17 clone 160_E_2 map 17, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
AC007952
AC007952
VERSION AC007952.1 GI:53323393
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 86408)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArillano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Headford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehocsky,J., Lieu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (02-JUL-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 515: contig of 515 bp in length
* 3281: contig of 2766 bp in length
* 516 3281: contig of 2766 bp in length
* 3282 5794: contig of 2513 bp in length
* 5795 8914: contig of 3120 bp in length
* 8915 27551: contig of 18637 bp in length
* 27552 39698: contig of 12147 bp in length
* gap of unknown length

FEATURES
source
1. .86408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC"
/map="17"
/chromosome="17"

BASE COUNT 21059 a 21431 c 23192 g 19868 t 858 others
ORIGIN

Query Match      77.0%; Score 15.4; DB 41; Length 86408;
Best Local Similarity 94.1%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cgtacctgacttagcc 20
|||||
Db 52441 CCATACCTGACTTAGCC 52457

RESULT 13
HSAJ9617/c
LOCUS
DEFINITION Homo sapiens chromosome 17 clone PAC RPCI-1 149D14 map 17pil.2, ***
SEQUENCING IN PROGRESS ***, 4 unordered pieces.
AC009617
AC009617.2 GI:6982187
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 128550)
Radelof,U., Hennig,S., Ramser,J., Francis,F., Steffens,C.,
Klein,M., Seranski,P., Poustka,A., Reinhardt,R. and Lehrach,H.
Unpublished
REFERENCE 2 (bases 1 to 128550)
MPIMG.
Direct Submission
Submitted (17-JUL-1998) MPIMG, Abt.Lehrach, Max Planck Institut
fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany
On Feb 16, 2000 this sequence version replaced gi:6729594.
HTGS_phase1: Unordered fragments separated by 800 N nucleotides
contig 01 1. .2165
contig 02 2965. .17224
contig 03 18025. .26787
contig 04 27588. .128550
at the Max-planck-Institute for Molecular Genetics.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2165: contig of 2165 bp in length
* 2966 2965: gap of 800 bp
* 17225 18024: contig of 14259 bp in length
* 18025 26787: contig of 800 bp
* 26788 27587: contig of 8763 bp in length
* 27588 128550: contig of 100963 bp in length.
Location/Qualifiers
1. .128550
/organism="Homo sapiens"
```

```

/db_xref="taxon:9606"
/chromosome="17"
/map="17p11.2"
/clone="PAC RPCI-1 149D14"
/clone_lib="RPCI1.3-5 Human PAC library, originating
institute: Roswell Park Cancer Institute, creator: Pieter
de Jong, P. Ioannou"
/note="region between markers D17S842-D17S953"
BASE COUNT 28413 a 34621 c 34489 g 28614 t 2413 others
ORIGIN

Query Match 77.0%; Score 15.4; DB 33; Length 128550;
Best Local Similarity 94.1%; Pred. No. 5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cagtaacctgacttagcc 20
||| ||||| ||||| |||
Db 3400 CCATACCTGACTTAGCC 3384

RESULT 14
AC010362
LOCUS Homo sapiens chromosome 5 clone CITB-HL_2037I18, WORKING DRAFT
DEFINITION AC010362.1 173219 bp DNA HTG 20-DEC-1999
SEQUENCE, 59 unordered pieces.
ACCESSION AC010362
VERSION AC010362.2 GI:6606052
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173219)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173219)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Dec 20, 1999 this sequence version replaced gi:5882572.
-----Genome Center
Center: Joint Genome Institute
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 173219; sum-of-contigs estimation
Estimated insert size: 130000; pulse field gel estimation
Quality coverage: 4.56x in Q20 bases; pulse field gel estimation
Quality coverage: 3.43x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 985: contig of 985 bp in length
* 886 gap of unknown length
* 2006: contig of 1021 bp in length
* 2007 gap of unknown length
* 3036: contig of 1030 bp in length
* 3037 gap of unknown length
* 4029: contig of 993 bp in length
* 4030 gap of unknown length
* 5462: contig of 1433 bp in length
* 5463 gap of unknown length
* 6443: contig of 981 bp in length
* 6444 gap of unknown length
* 7473: contig of 1030 bp in length

```

```

* 7474 9000: contig of 1527 bp in length
* gap of unknown length
* 9001 10051: contig of 1051 bp in length
* gap of unknown length
* 10052 11423: contig of 1372 bp in length
* gap of unknown length
* 11424 12580: contig of 1157 bp in length
* gap of unknown length
* 12581 13893: contig of 1313 bp in length
* gap of unknown length
* 13894 14893: contig of 1000 bp in length
* gap of unknown length
* 14894 16498: contig of 1605 bp in length
* gap of unknown length
* 16499 17824: contig of 1326 bp in length
* gap of unknown length
* 17825 19055: contig of 1231 bp in length
* gap of unknown length
* 19056 20282: contig of 1227 bp in length
* gap of unknown length
* 20283 21993: contig of 1711 bp in length
* gap of unknown length
* 21994 23858: contig of 1865 bp in length
* gap of unknown length
* 23859 25425: contig of 1567 bp in length
* gap of unknown length
* 25426 26664: contig of 1239 bp in length
* gap of unknown length
* 26665 28537: contig of 1873 bp in length
* gap of unknown length
* 28538 29750: contig of 1213 bp in length
* gap of unknown length
* 29751 31505: contig of 1754 bp in length
* gap of unknown length
* 31505 32519: contig of 1014 bp in length
* gap of unknown length
* 32519 33560: contig of 1042 bp in length
* gap of unknown length
* 33561 34885: contig of 1325 bp in length
* gap of unknown length
* 34886 35887: contig of 1001 bp in length
* gap of unknown length
* 35887 37375: contig of 1489 bp in length
* gap of unknown length
* 37375 38731: contig of 1356 bp in length
* gap of unknown length
* 38731 40917: contig of 2186 bp in length
* gap of unknown length
* 40917 42269: contig of 1352 bp in length
* gap of unknown length
* 42270 43940: contig of 1671 bp in length
* gap of unknown length
* 43941 45207: contig of 1267 bp in length
* gap of unknown length
* 45208 46466: contig of 1258 bp in length
* gap of unknown length
* 46466 47715: contig of 1250 bp in length
* gap of unknown length
* 47716 49034: contig of 1319 bp in length
* gap of unknown length
* 49035 50138: contig of 1104 bp in length
* gap of unknown length
* 50139 51181: contig of 1043 bp in length
* gap of unknown length
* 51182 52386: contig of 1205 bp in length
* gap of unknown length
* 52387 54456: contig of 2070 bp in length
* gap of unknown length
* 54457 55988: contig of 1532 bp in length
* gap of unknown length
* 55989 57021: contig of 1033 bp in length
* gap of unknown length

```

```
* 57022 58730: contig of 1709 bp in length
* gap of unknown length
* 58731 60058: contig of 1328 bp in length
* gap of unknown length
* 60059 61791: contig of 1733 bp in length
* gap of unknown length
* 61792 63374: contig of 1583 bp in length
* gap of unknown length
* 63375 64510: contig of 1136 bp in length
* gap of unknown length
* 64511 66265: contig of 1755 bp in length
* gap of unknown length
* 66266 69018: contig of 2753 bp in length
* gap of unknown length
* 69019 75389: contig of 6371 bp in length
* gap of unknown length
* 75390 80649: contig of 5260 bp in length
* gap of unknown length
* 80650 88197: contig of 7548 bp in length
* gap of unknown length
* 88198 94405: contig of 6208 bp in length
* gap of unknown length
* 94406 105418: contig of 11013 bp in length
* gap of unknown length
* 105419 115257: contig of 9839 bp in length
* gap of unknown length
* 115258 126202: contig of 10945 bp in length
* gap of unknown length
* 126203 141671: contig of 15469 bp in length
* gap of unknown length
* 141672 173219: contig of 31548 bp in length.
FEATURES
    source
    1..173219
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="5"
    /clone="CITB-H1_2037118"
BASE COUNT 53574 a 34847 c 36076 g 48558 t 164 others
ORIGIN

Query Match 77.0%; Score 15.4; DB 44; Length 173219;
Best Local Similarity 94.1%; Pred. No. 4.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ttgcgcgtactgactta 17
|||||||
Db 114775 TTGCCGTACTTGACTTA 114791

RESULT 15
AC012246 184541 bp DNA HTG 19-NOV-1999
LOCUS Homo sapiens clone RP11-16L19, WORKING DRAFT SEQUENCE, 8 unordered
DEFINITION pieces.
ACCESSION AC012246
VERSION AC012246.2 GI:6453959
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184541)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Farreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
```

```
TITLE
JOURNAL
COMMENT
On Nov 19, 1999 this sequence version replaced gi:6091798.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L3572
Center clone name: 16_L_19
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160082 bases at least Q40
Consensus quality: 175287 bases at least Q30
Consensus quality: 182134 bases at least Q20
Insert size: 188000; agarose-fp
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1029: contig of 1029 bp in length
* 1030 12093: gap of unknown length
* 12094 22982: contig of 11064 bp in length
* 22983 37256: gap of unknown length
* 37257 51513: contig of 10889 bp in length
* 51514 92507: gap of unknown length
* 92508 134662: contig of 14274 bp in length
* 134663 184541: gap of unknown length
* 184541: contig of 49879 bp in length.
FEATURES
    source
    1..184541
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-16L19"
    /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 45363 a 48606 c 46925 g 43643 t 4 others
ORIGIN

Query Match 77.0%; Score 15.4; DB 42; Length 184541;
Best Local Similarity 94.1%; Pred. No. 4.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ccgtacctgacttagcc 20
```

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
Morrow,P., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 19, 1999 this sequence version replaced gi:6091798.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submission@genome.wi.mit.edu  
----- Project Information  
Center project name: L3572  
Center clone name: 16\_L\_19  
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 160082 bases at least Q40  
Consensus quality: 175287 bases at least Q30  
Consensus quality: 182134 bases at least Q20  
Insert size: 188000; agarose-fp  
Quality coverage: 4.9 in Q20 bases; agarose-fp  
Quality coverage: 5.0 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1029: contig of 1029 bp in length  
\* 1030 12093: gap of unknown length  
\* 12094 22982: contig of 11064 bp in length  
\* 22983 37256: gap of unknown length  
\* 37257 51513: contig of 10889 bp in length  
\* 51514 92507: gap of unknown length  
\* 92508 134662: contig of 14274 bp in length  
\* 134663 184541: gap of unknown length  
\* 184541: contig of 49879 bp in length.

FEATURES  
 source  
 1..184541  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-16L19"  
 /clone\_lib="RPC1-11 Human Male BAC"  
BASE COUNT 45363 a 48606 c 46925 g 43643 t 4 others  
ORIGIN

Query Match 77.0%; Score 15.4; DB 42; Length 184541;  
Best Local Similarity 94.1%; Pred. No. 4.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 ccgtacctgacttagcc 20

|||||  
Db 154427 CCGTACCTGACTCAGCC 154443

Search completed: May 23, 2000, 09:40:01  
Job time: 8777 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 23, 2000, 09:42:02 ; Search time 88.1 Seconds  
(without alignments)  
56.797 Million cell updates/sec

Title: US-08-945-805-2

Perfect score: 20

Sequence: 1 ttgccgtacctgacttagcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	V59191	Random sequence ol
2	15.2	76.0	2520	1 V52359	Streptococcus pneu
3	15.2	76.0	4018	1 Q63879	Polyhydroxyalkanoa
4	14.8	74.0	16397	1 V74516	Staphylococcus aur
5	14.2	71.0	1394	1 X07408	Zea mays Ms45 male
6	14.2	71.0	1394	1 X07409	Zea mays Ms45 male
7	14.2	71.0	1512	1 V23482	Pseudomonas Xcpr s
8	14.2	71.0	1763	1 Q55316	Lactobacillus brev
9	14.2	71.0	2285	1 V80611	Kidney injury asso
10	14.2	71.0	3958	1 V74480	Staphylococcus aur
11	14.2	71.0	17612	1 V23494	Pseudomonas Xpc. O
12	14	70.0	1462	1 T35111	Enhanced expressio
13	14	70.0	1467	1 V82457	Triticum sp. cyste
14	13.8	69.0	224	1 T26337	Human gene signatu
15	13.8	69.0	395	1 T26460	Human gene signatu
16	13.8	69.0	711	1 V75115	Staphylococcus aur
17	13.8	69.0	1482	1 T26994	Serine threonine k
18	13.8	69.0	2938	1 X25656	Human endogenous r
19	13.8	69.0	3893	1 X13423	Enterococcus faeca
20	13.8	69.0	6605	1 V23545	Human gravin polyp
21	13.8	69.0	43280	1 T80413	Tyactone synthase
22	13.6	68.0	22	1 T31191	Friend virus strai
23	13.6	68.0	233	1 Q61007	Human brain Expres
24	13.6	68.0	236	1 V20258	Probe (159) for mi
25	13.6	68.0	238	1 Q41191	Human endo-exonucl
26	13.6	68.0	958	1 T96064	Apoptosis associat
27	13.6	68.0	1000	1 X20723	Polynucleotide seq
28	13.6	68.0	1689	1 T08083	Alpha-1,3/4-fucosi
29	13.6	68.0	2614	1 V84219	Mouse striated mus
30	13.6	68.0	2635	1 V84521	Human secreted pro
31	13.6	68.0	3185	1 Q24888	Plasmid contg. the
32	13.6	68.0	3185	1 Q99449	Saccharomyces olea
33	13.6	68.0	3185	1 Q99440	Saccharomyces olea
34	13.6	68.0	3442	1 V40528	Homo sapiens DL185

35 13.6 68.0 4071 1 Q04524 Alpha-1-6-glucanas  
36 13.6 68.0 6363 1 X28091 plasmid pACYC.pol.  
37 13.6 68.0 8323 1 Q58700 LTR clone of FB29  
38 13.6 68.0 8323 1 Q81042 Friend murine leuk  
39 13.6 68.0 8323 1 T80058 LTR clone of FB29  
c 40 13.6 68.0 10367 1 Q58701 Sequence of plasm  
c 41 13.6 68.0 10367 1 T80057 plRB332. Expressio  
c 42 13.6 68.0 25002 1 V52181 Streptococcus pneu  
43 13.6 68.0 80240 1 V83940 NC-contig derived  
44 13.6 68.0 80595 1 V83939 HC-contig derived  
45 13.6 68.0 110000 1 V30458\_3 continuation (4 of

#### ALIGNMENTS

##### RESULT 1

V59191  
ID V59191 standard; DNA; 20 BP.  
AC V59191;  
DT 21-DEC-1998 (first entry)  
DE Random sequence oligonucleotide.  
KW Bak gene; promoter; human; apoptosis; ischaemia; tumour; cancer;  
KW infection; HIV; neurodegenerative disorder; hair loss;  
KW inflammatory bowel disease; transplant rejection; therapy; Bcl-2;  
KW ISRE: interferon-gamma response element; GAS;  
KW gamma interferon-activated sequence; ds.  
OS Synthetic.  
PN Homo sapiens.  
PS WO9835659-A1.  
PD 20-AUG-1998.  
PF 17-FEB-1998; U02845.  
PR 18-FEB-1997; US-038412.  
PA (LXRB-) LXR BIOTECHNOLOGY INC.  
PI Kiefer MC, Ossina N;  
DR WPI; 98-506279/43.  
PT New isolated bak promoter - used for identifying agents that  
PT regulate the expression of a gene involved in apoptosis in cells for  
PT use in therapy  
PS Example 11; Page 36; 69pp; English.  
CC This double-stranded oligonucleotide with random sequence was used  
CC as a control in experiments to examine regulation of human Bak  
CC expression by an ISRE 'decoy' cis-element. Oligonucleotides  
CC corresponding to the bak promoter (see V59181) interferon-gamma  
CC response element (ISRE, see V59182), gamma-interferon activated  
CC sequence (GAS, see V59187) and NF-kappa 1 sequence (see V59189)  
CC were used as decoy elements to regulate Bak expression. GAS and  
CC NF-kappa B decoy DNAs failed to inhibit Bak expression. The  
CC ISRE decoy, but not the random-sequence decoy, markedly inhibited  
CC Bak expression in response to interferon-gamma. The ISRE  
CC cis-element is critical for Bak gene activation. ISRE decoy DNA  
CC can be used as a tool for inhibition of Bak transcription.  
SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T;

Query Match 100.0%; Score 20; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.046;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgccgtacctgacttagcc 20

Db 1 TTGCCGTACTGACTTAGCC 20

##### RESULT 2

V52359/c

ID V52359 standard; DNA; 2520 BP.

AC V52359;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:226.

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

PN WO9818931-A2.  
 PD 07-MAY-1998.  
 PF 30-OCT-1997; U19588.  
 PR 31-OCT-1996; US-029960.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,  
 PI Kunsch CA, Rosen CA;  
 DR WPI; 98-272225/24.  
 PT Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1; Page 1241-1243; 1409pp; English.  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridise to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 SQ Sequence 2520 BP; 646 A; 515 C; 543 G; 813 T;

Query Match 76.0%; Score 15.2; DB 1; Length 2520;  
 Best Local Similarity 85.0%; Pred. No. 28;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttgccgtacctgacttagcc 20  
 ||| | ||||| |||||  
 Db 308 TTGACATACCTGAGTTAGCC 289

RESULT 3  
 Q63879/C  
 ID Q63879 standard; DNA; 4018 BP.  
 AC Q63879;  
 DT 07-DEC-1994 (first entry)  
 DE Polyhydroxyalkanoate decomposing activity polypeptide.  
 KW Polyhydroxyalkanoate decomposing enzyme; PHA; ds.  
 OS Zoogloea ramigera.  
 FS Key Location/Qualifiers  
 FT cds 395..1153  
 FT /\*tag= a  
 FT /product= a  
 FT /note= "ORF 1"  
 FT 3173..3958  
 FT /\*tag= b  
 FT /product= b  
 FT /note= "ORF 2"  
 FT J06086681-A.  
 PN 29-MAR-1994.  
 PD 07-SEP-1992; 279099.  
 PR 07-SEP-1992; JP-279099.  
 PA (ELED ) DENKI KAGAKU KOGYO KK.  
 PA (UYKA-) GH KANAGAWA DAIGAKU.  
 PA (CHIK-) ZH CHIRYU KANKYO SANGYO GIJITSU KENKYU.  
 DR WPI; 94-140539/17.  
 DR P-PSDB; R53915 AND R63805.

PT DNA encoding poly-hydroxy-alkanoate decomposing enzyme - useful  
 PT for preparing poly-hydroxy-alkanoate in large quantities  
 PS Claim 3; Page 9-14; 15pp; Japanese.  
 CC Q63879 shows two ORFs, the products of which show  
 CC polyhydroxyalkanoate decomposing activity. The method used allows  
 CC the polypeptide to be prepared in large quantities.  
 SQ Sequence 4018 BP; 767 A; 1298 C; 1305 G; 648 T;

Query Match 76.0%; Score 15.2; DB 1; Length 4018;  
 Best Local Similarity 85.0%; Pred. No. 29;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttgccgtacctgacttagcc 20  
 ||||| ||||| ||||| |||||  
 Db 3345 TTGCCGTTCCGCTTCGCC 3326

RESULT 4  
 V74516/C  
 ID V74516 standard; DNA; 16397 BP.  
 AC V74516;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #205.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FS Key Location/Qualifiers  
 FT misc\_feature 421..480  
 FT /\*tag= a  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 2221..2280  
 FT /\*tag= b  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 4021..4080  
 FT /\*tag= c  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 5821..5880  
 FT /\*tag= d  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 7621..7680  
 FT /\*tag= e  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 9421..9480  
 FT /\*tag= f  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 11221..11280  
 FT /\*tag= g  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 13021..13080  
 FT /\*tag= h  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"

```
FT FT /*tag= h
FT FT /note= "these bases represent a line of missing text in
FT FT the sequence listing in the specification. They
FT FT are included to maintain the nucleotide numbering
FT FT given in the specification for this DNA sequence".
FT FT 14821. 14880
FT FT misc_feature 14821. 14880
FT FT /*tag= 1
FT FT /note= "these bases represent a line of missing text in
FT FT the sequence listing in the specification. They
FT FT are included to maintain the nucleotide numbering
FT FT given in the specification for this DNA sequence".
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S aureus vaccines
PS Claim 1; Page 922-931; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 16397 BP; 4397 A; 3179 C; 2525 G; 5744 T;

Query Match 74.0%; Score 14.8; DB 1; Length 16397;
Best Local Similarity 88.9%; Pred. No. 55;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttagc 19
||||| ||| |||||
Db 7753 TGCCGTACTGAAATGAC 7736

RESULT 5
X07408/c
ID X07408 standard; DNA; 1394 BP.
AC X07408;
DT 08-JUN-1999 (first entry)
DE Zea mays Ms45 male tissue-preferred regulatory region.
KW Ms45; male; tissue-preferred; regulatory region; plant cells;
KW plant tissue; differentiated; maize; hybrid seed; fertility; ss.
OS Zea mays.
PN W09859061-A1.
PD 30-DEC-1998.
PF 19-JUN-1998; U12895.
PR 23-JUN-1997; US-880499.
PA (PION-) PIONEER HI-BRED INT INC.
PI Albertsen MC, Fox TW, Garnaat CW, Huffman GA, Kendall TL;
DR WPI; 99-105628/09.
PT New nucleic acid encoding a Ms45 male tissue-preferred regulatory
PT region - useful in mediating plant fertility, especially hybrid seed
PT production
PS Claim 2; Pages 22-23; 39pp; English.
```

```
CC The sequence is that encoding an Ms45 male tissue-preferred
CC regulatory region. It may be used in the construction of a vector
CC for a method of producing exogenous genes in a male tissue-
CC preferred manner, which is useful in restoring or conferring
CC fertility, such as in hybrid seed production. In conferring
CC fertility, a monocot/dicot plant is transformed with the
CC exogenous nucleotide sequence (a male sterility gene, preferably
CC Ms45), which encodes a product selected from auxins, rolB and
CC diphtheria toxin. Hybrid seeds are produced by cross-pollinating
CC maize male fertile and infertile plants. 231 G; 441 T;
SQ Sequence 1394 BP; 411 A; 311 C; 232 G; 442 T;

Query Match 71.0%; Score 14.2; DB 1; Length 1394;
Best Local Similarity 84.2%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttagcc 20
||||| ||||| ||| |
Db 1149 TGCCGAACCTGACGTAGGC 1131

RESULT 6
X07409/c
ID X07409 standard; DNA; 1394 BP.
AC X07409;
DT 08-JUN-1999 (first entry)
DE Zea mays Ms45 male tissue-preferred regulatory region.
KW Ms45; male; tissue-preferred; regulatory region; plant cells;
KW plant tissue; differentiated; hybrid seed; fertility; ss.
OS Zea mays.
PN W09859061-A1.
PD 30-DEC-1998.
PF 19-JUN-1998; U12895.
PR 23-JUN-1997; US-880499.
PA (PION-) PIONEER HI-BRED INT INC.
PI Albertsen MC, Fox TW, Garnaat CW, Huffman GA, Kendall TL;
DR WPI; 99-105628/09.
PT New nucleic acid encoding a Ms45 male tissue-preferred regulatory
PT region - useful in mediating plant fertility, especially hybrid seed
PT production
PS Claim 3; Pages 23-24; 39pp; English.
CC The sequence is that encoding an Ms45 male tissue-preferred
CC regulatory region. It may be used in the construction of a vector
CC for a method of producing exogenous genes in a male tissue-
CC preferred manner, which is useful in restoring or conferring
CC fertility, such as in hybrid seed production. In conferring
CC fertility, a monocot/dicot plant is transformed with the
CC exogenous nucleotide sequence (a male sterility gene, preferably
CC Ms45), which encodes a product selected from auxins, rolB and
CC diphtheria toxin. Hybrid seeds are produced by cross-pollinating
CC maize male fertile and infertile plants. 232 G; 442 T;
SQ Sequence 1394 BP; 411 A; 309 C; 232 G; 442 T;

Query Match 71.0%; Score 14.2; DB 1; Length 1394;
Best Local Similarity 84.2%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttagcc 20
||||| ||||| ||| |
Db 1149 TGCCGAACCTGACGTAGGC 1131

RESULT 7
V23482
ID V23482 standard; DNA; 1512 BP.
AC V23482;
DT 24-JUL-1998 (first entry)
DE Pseudomonas xcpR secretion factor coding sequence.
KW Kinase; LipQ; LipR; lipase expression regulator; DNA binding regulator;
KW sigma 54 promoter; secretion factor; lux-box binding element;
KW orfV-box binding element; regulation cascade; ss.
```

OS Pseudomonas alcaligenes.  
 PN WO9806836-A2.  
 PD 19-FEB-1998.  
 PR 15-AUG-1997; U14450.  
 PR 16-AUG-1996; US-699092.  
 PA (GEMV ) GENENCOR INT INC.  
 PI Gerritse G, Quax WJ;  
 DR WPI: 98-159528/14.  
 DR P-PSDB; W53830.  
 PT Nucleic acids encoding proteins involved in the lipase regulation and  
 PT cascade from P. alcaligenes - useful for controlling production and  
 PT secretion of heterologous proteins in P. alcaligenes  
 PS Claim 13; Page 25; 106pp; English.  
 CC This sequence encodes the XcpR secretion factor of Pseudomonas  
 CC alcaligenes. The DNA represents a nucleic acid of the invention. The  
 CC nucleic acid encodes: (a) a kinase from a Pseudomonad that regulates the  
 CC expression of a lipase; (b) a DNA binding regulator from a Pseudomonad  
 CC that regulates the expression of a lipase; (c) a Pseudomonas alcaligenes  
 CC upstream activating sequence; (d) a P. alcaligenes sigma 54 promoter that  
 CC regulates the expression of a lipase; (e) a P. alcaligenes secretion  
 CC factor selected from XcpP, O, R, S, T, U, V, W, X, Y and Z, and OrfV, X,  
 CC Y; (f) a P. alcaligenes lux-box binding element; and (g) a orfV-box  
 CC binding element. The nucleic acids represent parts of a regulation  
 CC cascade, comprising at its heart a kinase and a DNA binding regulator.  
 CC These sequences can be used for the production of heterologous proteins  
 CC in a host cell. The cascade also comprises secretion factors which can  
 CC enhance the secretion of produced proteins.  
 SQ Sequence 1512 BP; 238 A; 566 C; 489 G; 219 T;

Query Match 71.0%; Score 14.2; DB 1; Length 1512;  
 Best Local Similarity 84.2%; Pred. No. 89;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgcgcgtacctgacttagcc 20  
 ||||| ||||| |||||  
 Db 288 TGCAGAACTTGACCTAGCC 306

RESULT 8  
 Q55316/c  
 ID Q55316 standard; DNA; 1763 BP.  
 AC Q55316;  
 DT 08-JUL-1994 (first entry)  
 DE Lactobacillus brevis DSM20556 surface protein gene.  
 KW Lactobacillus brevis; primer: surface protein; fusion protein;  
 KW antisense; repression; transcription; translation; expression;  
 KW regulatory elements; ss.  
 OS Lactobacillus brevis.  
 FH Key Location/Qualifiers  
 FT cds 268..1662  
 FT /tag= a  
 FT /product= Lactobacillus brevis surface protein.

PN WO9400581-A.  
 PD 06-JAN-1994.  
 PR 24-JUN-1993; FI0273.  
 PR 30-JUN-1992; US-906320.  
 PA (VIAG-) VIAGEN OY.  
 PI Palva AM, Palva IA;  
 DR WPI: 94-026217/03.  
 DR P-PSDB; R47972.  
 PT Lactobacillus expression system using surface protein gene  
 PT sequences - using lactobacillus coat protein expression and  
 PT secretory regulatory elements  
 PS Disclosure; Figure 2B; 85pp; English.  
 CC A recombinant vector comprising at least one Lactobacillus coat  
 CC protein expression regulatory element can be used to express a  
 CC heterologous desired sequence. The regulatory element is selected  
 CC from a coat protein promoter, a full length coat protein secretion  
 CC signal sequence, and one or more domains of a coat protein secretion  
 CC sequence. Desired proteins can be expressed as part of a  
 CC fusion protein. Alternatively, the surface protein (SP) gene  
 CC regulatory elements may be used to direct the synthesis of antisense

CC RNA in a host cell, to inhibit or repress transcription/translation  
 CC of an endogenous protein gene or mRNA. This may be desired when  
 CC expressing a SP fusion protein or other construct on the host  
 CC surface in place of the native SP protein.  
 SQ Sequence 1763 BP; 489 A; 361 C; 378 G; 535 T;

Query Match 71.0%; Score 14.2; DB 1; Length 1763;  
 Best Local Similarity 84.2%; Pred. No. 91;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcgcgtacctgacttagc 19  
 ||||| ||||| |||||  
 Db 994 TTGCTGTAACTGACTTATC 976

RESULT 9  
 V80611/c  
 ID V80611 standard; cDNA; 2285 BP.  
 AC V80611;  
 DT 01-MAR-1999 (first entry)  
 DE Kidney injury associated molecule HW059 cDNA clone.  
 KW Kidney injury associated molecule; kidney injury related molecule;  
 KW KIM; tissue growth promotion; regeneration; renal condition;  
 KW acute renal failure; acute nephritis; tumour; ds.  
 OS Rattus sp.  
 FH Key Location/Qualifiers  
 FT cds 9..1274  
 FT /tag= a  
 FT /product= "Kidney injury associated molecule"  
 FT misc\_feature 191..2162  
 FT /tag= b  
 FT /label= SAC\_24457  
 PN WO953071-A1.  
 PD 26-NOV-1998.  
 PF 22-MAY-1998; U10547.  
 PR 23-MAY-1997; US-047491.  
 PR 23-MAY-1997; US-047490.  
 PA (BIOJ ) BIOGEN INC.  
 PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;  
 DR WPI: 99-045312/04.  
 DR P-PSDB; W86326.  
 PT Kidney injury-associated molecule, KIM, polypeptides - upregulated  
 PT in injured or regenerating tissues, useful to promote tissue growth  
 PT and regeneration, especially to treat renal conditions  
 PS Claim 9; Page 133-135; 213pp; English.  
 CC The present sequence represents a kidney injury associated molecule  
 CC (KIM) cDNA clone. KIM proteins can be administered therapeutically  
 CC by expressing KIM encoding polynucleotides, to promote growth and/or  
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
 CC are upregulated in injured or regenerating (especially renal) tissues.  
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
 CC therapeutically, e.g. these or the KIM proteins may be included with an  
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
 CC prophylaxis of conditions associated with dysfunction/disregulation of  
 CC KIM genes or proteins, especially renal diseases or impairments of renal  
 CC function in humans (e.g. acute renal failure, acute nephritis). The  
 CC polynucleotides can be used to produce antisense sequences which, when  
 CC internalised into cells, can disrupt expression of a cellular KIM gene,  
 CC also useful in therapy (e.g. to block the growth of tumours dependent on  
 CC KIM for growth) or compositions. The proteins and polynucleotides are  
 CC useful diagnostically e.g. to detect and quantify renal injury/disease  
 CC (indicative of increased risk, or presence of, renal injury or impaired  
 CC function), or abnormal responses to tissue injury (indicative of  
 CC increased risk, or presence of, an autoimmune response or abnormal  
 CC tissue growth arising from/affecting renal tissue). The proteins can  
 CC also be used to locate KIM-producing cells (especially specific loci,  
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
 CC arising from/affecting renal tissue), by contacting cells with an  
 CC imageable KIM-binding reagent and imaging reagent accumulation.  
 SQ Sequence 2285 BP; 454 A; 656 C; 689 G; 486 T;

Query Match	71.0%;	Score 14.2;	DB 1;	Length 2285;
Best Local Similarity	84.2%;	Pred. No. 93;		
Matches 16;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0;

Qy 2 tgcgtaactgacttagcc 20  
 ||||| | |||||  
 Db 1211 TGCCGTACAGGCCTTAGCC 1193

```

RESULT 10
V74480
ID V74480 standard; DNA; 3958 BP.
AC V74480;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #169.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 361..420
FT /*tag= a
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence."
FT misc_feature 2161..2220
FT /*tag= b
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence."
FT misc_feature

```

EP-786519-A2.  
30-JUL-1997.  
07-JAN-1997; 100117.  
05-JAN-1996; US-009861.  
(HUMA-) HUMAN GENOME SCI INC.  
Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
Rosen CA,  
WPI; 97-374922/35.  
Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
stored on computer readable medium and used in the production of  
anti-S.aureus vaccines  
Claim 1; Page 827-829; 3271pp; English.  
This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
of the invention. The DNA sequences are recorded on a computer readable  
medium, preferably selected from a floppy or hard disk, random access  
memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
the S.aureus DNA sequences allows putative functions to be assigned so  
that protein-encoding or regulatory regions of commercial, therapeutic or  
industrial importance can be obtained. Specifically, sequences which are  
likely to encode antigens have been identified and these polypeptides can  
be used in a vaccine composition against S.aureus infection. The  
polypeptides can also be used in a kit for the immunodetection of  
S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
skin and surgical wound infections, scalded skin syndrome, toxic shock  
syndrome, etc. Organisms transformed with the DNA sequences can be used  
for recombinant production of the polypeptides. The new DNA sequences  
(and their fragments) are useful as primers or probes for isolating  
homologues of any of the S.aureus DNA sequences contained on the  
computer readable medium.  
Sequence 3958 BP; 1470 A; 537 C; 753 G; 1074 T;

Qy	1	ttgccgtac	ctgacttagc	19
Db	3059	TTGCTGTAC	CTGCCTTAAC	3077

RESULT 11

V23494 V23494 standard; DNA; 17612 BP.

AC V23494;

DT 24-JUL-1998 (first entry)

DE Pseudomonas Xpc, OrfY, OrfV, OrfX operon.

KW Kinase; LipQ; LipR; lipase expression regulator; DNA binding regulator;

KW sigma 54 promoter; secretion factor; lux-box binding element;

KW orfV-box binding element; regulation cascade; ss.

OS Pseudomonas alcaligenes.

EH Key Location/Qualifiers

CD5 1..1176

FT /\*tag= a

FT /product= OrfY

FT complement (3513..1564)

FT /\*tag= b

FT /product= xcpQ

FT complement (4151..3510)

FT /\*tag= c

FT /product= XcpP

FT 4416..7049

FT /\*tag= d

FT /product= OrfV

FT 7168..9909

FT /\*tag= e

FT /product= OrfX

FT 10122..11633

FT /\*tag= f

FT /product= xcpR

FT 11633..12847

FT /\*tag= g

FT /product= Xcps

FT 12853..13275

FT /\*tag= h

FT /product= xcpT

FT 13282..13806

FT /\*tag= i

FT /product= xcpU

FT 13803..14192

FT /\*tag= j

FT /product= xcpV

FT 14189..14872

FT /\*tag= k

FT /product= xcpW

FT 14869..15822

FT /\*tag= l

FT /product= xcpX

FT 15819..16964

FT /\*tag= m

FT /product= xcpY

FT 16961..17473

FT /\*tag= n

FT /product= xcpZ

PN W09806836-A2.

PD 19-FEB-1998.

PF 15-AUG-1997; U14450

PR 16-AUG-1996; US-699092.

PA (GENV ) GENENCOR INT INC.

PI Gerritse G, Quax WJ;

PI WPI; 98-159528/14.

DR P-PSDB; W99999.

DR Nucleic acids encoding proteins involved in the lipase regulation

PT cascade from P. alcaligenes - useful for controlling production and

PT secretion of heterologous proteins in P. alcaligenes

PT Disclosure; Page 43-48; 106pp; English.

PS This sequence encodes the OrfV, OrfV, OrfX, and xpc secretion factors of

CC Pseudomonas alcaligenes. The DNA represents a nucleic acid of the

CC invention. The nucleic acids encode: (a) a kinase from a Pseudomonad that

CC regulates the expression of a lipase; (b) a DNA binding regulator from a

CC Pseudomonad that regulates the expression of a lipase; (c) a Pseudomonas

CC alcaligenes upstream activating sequence; (d) a P. alcaligenes sigma 54

CC promoter that regulates the expression of a lipase; (e) a P. alcaligenes  
 CC serfyn factor selected from xcpP, Q, R, S, T, U, V, W, X, Y and Z, and  
 CC OrfV, X, Y; (f) a P. alcaligenes lux-box binding element; and (g) a  
 CC orfV-box binding element. The nucleic acids represent parts of a  
 CC regulation cascade, comprising at its heart a kinase and a DNA binding  
 CC regulator. These sequences can be used for the production of heterologous  
 CC proteins in a host cell. The cascade also comprises secretion factors  
 CC which can enhance the secretion of produced proteins.  
 SQ Sequence 17612 BP; 2698 A; 6026 C; 6039 G; 2836 T;

Query Match 71.0%; Score 14.2; DB 1; Length 17612;  
 Best Local Similarity 84.2%; Pred. No. 1.2e+02;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgcctgacttgacttgacc 20  
 ||||| || |||| |||||  
 Db 10409 TGCCGAACCTTGACCTAGCC 10427

RESULT 12  
 T35111/c  
 ID T35111 standard; cDNA; 1442 BP.  
 AC T35111;  
 DT 21-NOV-1996 (first entry)  
 DE Enhanced expression senescence clone, SEEL.  
 KW Senescence related gene; expression; inhibition; acceleration; tomato;  
 KW lettuce; cabbage; banana; strawberry; wheat; maize; rice; rape; soybean;  
 KW delayed senescence; yield; protein content; quality; tolerance;  
 KW increased senescence; desiccant; ss.  
 OS Zea mays.  
 FH Key Location/Qualifiers  
 FT CDS 78..1160  
 FT /\*tag= a

PN W09507993-A1.  
 PD 23-MAR-1995.  
 PF 13-SEP-1994; G01990.  
 PR 13-SEP-1993; GB-018927.  
 PA (ZENE ) ZENECA LTD.  
 PI Drake CR, Farrell A, Grierson D, Hosken SE, John I;  
 PI Schuch WM, Smart CM, Thomas H;  
 PI WPI; 95-131361/17.  
 DR DNA constructs which modify expression of senescence-related genes -  
 PT useful to accelerate or inhibit senescence in crop plants  
 PS Claim 16; Page 55-56; 78pp; English.

CC The sequences given in T35095-133 are senescence related genes and  
 CC fragments which were used in the construct of the invention which  
 CC modifies the expression of at least one senescence related gene.  
 CC Using these constructs senescence may be inhibited or accelerated  
 CC in plants including tomato, lettuce, cabbage, banana, strawberry,  
 CC wheat, maize, rice, rape or soybean. Delayed senescence may  
 CC indirectly prolong the life of the plant, increase yield, increase  
 CC protein content of fruits, improve quality of leafy vegetables,  
 CC improve tolerance to disease, drought or other stress. Increased  
 CC senescence may more rapidly break down unwanted plant material and  
 CC so avoid the use of desiccants on crops. This sequence represents the  
 CC senescence-enhanced clone, SEEL, which is also known as clone  
 CC p16.4. SEEL is a cDNA of approx. 1.7 kb encoding a mRNA of approx.  
 CC 1.2 kb. The mRNA encoded by SEEL increases in abundance during maize  
 CC leaf senescence. SEEL shows homology to genes for two thiol proteases,  
 CC oryzain gamma from rice and aleurain from barley. The clone SEEL may  
 CC therefore encode a protease. This is supported by the finding that a 1  
 CC kb DNA fragment is amplified by PCR when one of the primers used is  
 CC derived from a region which is conserved in a range of thiol proteases.  
 CC The predicted amino acid sequence of the protein encoded by SEEL  
 CC suggests that the protein is 360 amino acids long with a mol. wt. of 39  
 CC kd. There are conserved motifs in the amino acid sequence, a putative  
 CC vacuolar signal, a Cys active site, a His active site and an Asn active  
 CC site. The predominantly hydrophilic protein sequence suggests that the  
 CC protein is soluble. SEEL is deposited as NCIMB 40382.  
 CC Sequence 1442 BP; 314 A; 408 C; 393 G; 327 T;

Query Match 70.0%; Score 14; DB 1; Length 1442;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcctgacttgacc 14  
 ||||| ||||| |||||  
 Db 280 TTGCCGTACCTGAC 267

RESULT 13  
 V82457/c  
 ID V82457 standard; DNA; 1467 BP.  
 AC V82457;  
 DT 16-MAR-1999 (first entry)  
 DE Triticum sp. cysteine proteinase #2 encoding DNA.  
 KW Triticum; wheat seed; cysteine proteinase; gluten; baking; ds.  
 OS Triticum sp.  
 FH Key Location/Qualifiers  
 FT CDS 47..1144  
 FT /\*tag= a

PN J10327886-A.  
 PD 15-DEC-1998.  
 PF 27-MAR-1998; 098140.  
 PR 31-MAR-1997; JP-114946.  
 PA (SHOS ) SHOWA SANGYO CO.  
 DR WPI; 99-109255/10.  
 DR P-PSDB; W89557.  
 PT New DNA coding cysteine proteinase originating from wheat seed -  
 PT useful for improving gluten for use in bakery process  
 PS Claim 2; Page 16-18; 29pp; Japanese.  
 CC The present sequence encodes a cysteine proteinase isolated from wheat  
 CC seed (Triticum sp.). The cysteine proteinase is useful for improving  
 CC gluten for use in the bakery process.  
 CC Sequence 1467 BP; 348 A; 423 C; 389 G; 305 T;

Query Match 70.0%; Score 14; DB 1; Length 1467;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcctgacttgacc 14  
 ||||| ||||| |||||  
 Db 264 TTGCCGTACCTGAC 251

RESULT 14  
 T26337/c  
 ID T26337 standard; cDNA to mRNA; 224 BP.  
 AC T26337;  
 DT 16-OCT-1996 (first entry)  
 DE Human gene signature HUMGS08577.  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 OS Homo sapiens.  
 PN W09514772-A1.  
 PD 01-JUN-1995.  
 PF 11-NOV-1994; J01916.  
 PR 12-NOV-1993; JP-355504.  
 PA (MATS/) MATSUBARA K.  
 PA (OKUB/) OKUBO K.  
 PI Matsubara K, Okubo K;  
 DR WPI; 95-206931/27.  
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1; Page 2060; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 224 BP; 66 A; 42 C; 43 G; 68 T;

Query Match 69.0%; Score 13.8; DB 1; Length 224;  
Best Local Similarity 88.2%; Pred. No. 1.2e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ttgcgcgtacctgactta 17  
||||| || |||||  
Db 126 TTGCCTTAGCTGACTTA 110

## RESULT 15

T26460/c  
ID T26460 standard; cDNA to mRNA; 395 BP.  
AC T26460;  
DT 23-OCT-1996 (first entry)  
DE Human gene signature HUMGS08702.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-353504.  
PA (MATSU) MATSUBARA K.  
PA (OKUBO) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 2090-2091; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 395 BP; 104 A; 67 C; 83 G; 125 T;

Query Match 69.0%; Score 13.8; DB 1; Length 395;  
Best Local Similarity 88.2%; Pred. No. 1.3e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ttgcgcgtacctgactta 17  
||||| || |||||  
Db 223 TTGCCTTAGCTGACTTA 207

Search completed: May 23, 2000, 09:42:07  
Job time: 8839 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:21:39 ; Search time 1428.99 Seconds  
(without alignments)

56.729 Million cell updates/sec

Title: US-08-945-805-2  
Perfect score: 20  
Sequence: 1 ttgccgtacctgacttagcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result      %      Query

No.	Score	Match	Length	DB	ID	Description
c 1	16.8	84.0	199	74	AV392961	AV392961 AV392961
c 2	16.8	84.0	541	74	AV397286	AV397286 AV397286
c 3	16.8	84.0	559	74	AV393452	AV393452 AV393452
c 4	16.8	84.0	583	74	AV387393	AV387393 AV387393
c 5	15.8	79.0	323	81	AW448957	UI-H-BI3-
c 6	15.8	79.0	401	92	AQ090391	GSSTC0101
c 7	15.8	79.0	421	28	CI8763	CI8763 CI8763 Huma
c 8	15.8	79.0	574	80	AW330086	TENU4832
c 9	15.8	79.0	579	40	AA911667	ol49f04.s
c 10	15.8	79.0	599	40	AA929849	vs5c09.r
c 11	15.4	77.0	605	107	AQ473994	CITBI-EI-
c 12	15.4	77.0	710	108	AQ528466	RPCI-11-3
c 13	15.2	76.0	166	46	AA957921	UI-R-EI-f
c 14	15.2	76.0	183	84	B71840	RPCI11-11C6
c 15	15.2	76.0	243	41	AI045089	UI-R-CL-k
c 16	15.2	76.0	275	51	AI714192	UI-R-AGJ-
c 17	15.2	76.0	311	46	AA963041	UI-R-EI-f
c 18	15.2	76.0	314	51	AI712948	UI-R-Y0-a
c 19	15.2	76.0	352	102	AQ104447	HS_3127_A
c 20	15.2	76.0	333	60	AI796378	WJ17a11.x
c 21	15.2	76.0	373	108	AQ547022	CITBI-EI-
c 22	15.2	76.0	381	109	AQ641025	RPCI93-Dp
c 23	15.2	76.0	391	23	H31184	EST104968 R
c 24	15.2	76.0	391	40	AA975297	og36g11.s
c 25	15.2	76.0	406	45	AI329823	b8f10ne.r
c 26	15.2	76.0	409	39	AA848316	EST191076
c 27	15.2	76.0	410	41	AI008388	EST202839
c 28	15.2	76.0	436	48	AI577025	UI-R-ABO-
c 29	15.2	76.0	460	21	T80557	Yd06a09.r1
c 30	15.2	76.0	460	41	AI012769	EST207220
c 31	15.2	76.0	462	30	AA241078	mv23609.r
c 32	15.2	76.0	472	92	AQ940604	Sheared D
c 33	15.2	76.0	477	41	AI008506	EST202957
c 34	15.2	76.0	479	102	AQ133308	HS_3043_A
c 35	15.2	76.0	480	106	AQ355443	CITBI-EI-
c 36	15.2	76.0	481	46	AI410534	EST238827
c 37	15.2	76.0	482	41	AI009314	EST203765
c 38	15.2	76.0	493	90	AQ818005	HS_5253_B
c 39	15.2	76.0	512	109	AQ645765	RPCI93-Dp
c 40	15.2	76.0	525	43	AI229667	EST226362
c 41	15.2	76.0	615	52	AI727243	BNLGH1755
c 42	15.2	76.0	636	92	AQ949013	Sheared D
c 43	15.2	76.0	653	69	AV382333	AV382333
c 44	15	75.0	555	93	AQ988317	RPCI-23-3
c 45	15	75.0	591	92	AQ920076	RPCI-23-2

## ALIGNMENTS

RESULT 1	
AV392961/c	
LOCUS	AV392961 199 bp mRNA EST 21-JAN-2000
DEFINITION	AV392961 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
DESCRIPTION	cDNA clone CM093c05_r, mRNA sequence.
ACCESSION	AV392961
VERSION	AV392961.1 GI:6547177
KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
AUTHORS	Chlamydomonadaceae; Chlamydomonas.
TITLE	1 (bases 1 to 199)
ASAMIZU, E., NAKAMURA, Y., SATO, S., FUKUZAWA, H. and TABATA, S.	
A Large Scale Structural Analysis of cDNAs in a Unicellular Green	
Alga, Chlamydomonas reinhardtii. I. Generation of 3433	
Non-redundant Expressed Sequence Tags	
DNA Res. 6, 369-373 (1999)	
On Jul 8, 1999 this sequence version replaced gi:5422458.	
Contact: Yasukazu Nakamura	
The First Laboratory for Plant Gene Research	
FEATURES	
source	1. .541
/organism="Chlamydomonas reinhardtii"	
/strain="C9"	
/db_xref="taxon:3055"	
/clone="CL77h05_r"	
/clone_lib="Chlamydomonas reinhardtii C9"	
/dev_stage="photoautotrophic growth"	
/note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	116 a 150 c 176 g 99 t
ORIGIN	
Query Match	84.0%; Score 16.8; DB 74; Length 541;
Best Local Similarity	90.0%; Pred. No. 52;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 ttgcgcgtacctgacttagcc 20
Db	93 TTGCCGTACCAGACTTAGCC 74

FEATURES	
source	1. .199
/organism="Chlamydomonas reinhardtii"	
/strain="C9"	
/db_xref="taxon:3055"	
/clone="CM093c05_r"	
/clone_lib="Chlamydomonas reinhardtii C9"	
/dev_stage="photoautotrophic growth"	
/note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	41 a 62 c 53 g 43 t
ORIGIN	
Query Match	84.0%; Score 16.8; DB 74; Length 199;
Best Local Similarity	90.0%; Pred. No. 43;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 ttgcgcgtacctgacttagcc 20
Db	103 TTGCCGTACCAGACTTAGCC 84
RESULT 2	
AV397286/c	
LOCUS	AV397286 541 bp mRNA EST 21-JAN-2000
DEFINITION	AV397286 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
DESCRIPTION	cDNA clone CL77h05_r, mRNA sequence.
ACCESSION	AV397286
VERSION	AV397286.1 GI:6551502
KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
AUTHORS	Chlamydomonadaceae; Chlamydomonas.
TITLE	1 (bases 1 to 541)
ASAMIZU, E., NAKAMURA, Y., SATO, S., FUKUZAWA, H. and TABATA, S.	
A Large Scale Structural Analysis of cDNAs in a Unicellular Green	
Alga, Chlamydomonas reinhardtii. I. Generation of 3433	
Non-redundant Expressed Sequence Tags	
DNA Res. 6, 369-373 (1999)	
On Mar 16, 1998 this sequence version replaced gi:2961775.	
Contact: Yasukazu Nakamura	
The First Laboratory for Plant Gene Research	
Kazusa DNA Research Institute	
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan	
Email: ynakamuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.	
FEATURES	
source	1. .541
/organism="Chlamydomonas reinhardtii"	
/strain="C9"	
/db_xref="taxon:3055"	
/clone="CL77h05_r"	
/clone_lib="Chlamydomonas reinhardtii C9"	
/dev_stage="photoautotrophic growth"	
/note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	116 a 150 c 176 g 99 t
ORIGIN	
Query Match	84.0%; Score 16.8; DB 74; Length 541;
Best Local Similarity	90.0%; Pred. No. 52;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 ttgcgcgtacctgacttagcc 20
Db	93 TTGCCGTACCAGACTTAGCC 74

```

RESULT 3
AV393452/c
LOCUS      559 bp      mRNA      EST      21-JAN-2000
DEFINITION Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
            CDNA clone CL07c08_r, mRNA sequence.
ACCESSION  AV393452
VERSION     AV393452.1  GI:6547668
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii.
            Chlamydomonas reinhardtii
ORGANISM   Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE  1 (bases 1 to 559)
AUTHORS   Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE     A Large Scale Structural Analysis of cDNAs in a Unicellular Green
            Alga, Chlamydomonas reinhardtii. I. Generation of 3433
            Non-redundant Expressed Sequence Tags
JOURNAL   DNA Res. 6, 369-373 (1999)
COMMENT   On Jul 9, 1999 this sequence version replaced gi:5434199.
            Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
            source
            1..559
            /organism="Chlamydomonas reinhardtii"
            /strain="C9"
            /db_xref="taxon:3055"
            /clone="CL07c08_r"
            /dev_stage="photoautotrophic growth"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT  131 a 158 c 176 g 94 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 74; Length 559;
Best Local Similarity 90.0%; Pred. No. 52;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 ttgcgcgtacctgacttagcc 20
      ||||| 1 |||||
Db  55 TTGCGGTAGCAGACTTAGCC 36

RESULT 4
AV387393/c
LOCUS      583 bp      mRNA      EST      21-JAN-2000
DEFINITION Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
            CDNA clone CM014g06_r, mRNA sequence.
ACCESSION  AV387393
VERSION     AV387393.1  GI:6541609
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii.
            Chlamydomonas reinhardtii
ORGANISM   Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE  1 (bases 1 to 583)
AUTHORS   Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE     A Large Scale Structural Analysis of cDNAs in a Unicellular Green
            Alga, Chlamydomonas reinhardtii. I. Generation of 3433
            Non-redundant Expressed Sequence Tags
JOURNAL   DNA Res. 6, 369-373 (1999)
COMMENT   On Mar 10, 1998 this sequence version replaced gi:2948559.
            Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
            source
            1..583
            /organism="Chlamydomonas reinhardtii"
            /strain="C9"
            /db_xref="taxon:3055"
            /clone="CM014g06_r"
            /dev_stage="photoautotrophic growth"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT  131 a 158 c 176 g 94 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 74; Length 559;
Best Local Similarity 90.0%; Pred. No. 52;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 ttgcgcgtacctgacttagcc 20
      ||||| 1 |||||
Db  55 TTGCGGTAGCAGACTTAGCC 36

RESULT 4
AV387393/c
LOCUS      583 bp      mRNA      EST      21-JAN-2000
DEFINITION Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
            CDNA clone CM014g06_r, mRNA sequence.
ACCESSION  AV387393
VERSION     AV387393.1  GI:6541609
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii.
            Chlamydomonas reinhardtii
ORGANISM   Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE  1 (bases 1 to 583)
AUTHORS   Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE     A Large Scale Structural Analysis of cDNAs in a Unicellular Green
            Alga, Chlamydomonas reinhardtii. I. Generation of 3433
            Non-redundant Expressed Sequence Tags
JOURNAL   DNA Res. 6, 369-373 (1999)
COMMENT   On Mar 10, 1998 this sequence version replaced gi:2948559.
            Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
            source
            1..583
            /organism="Chlamydomonas reinhardtii"
            /strain="C9"
            /db_xref="taxon:3055"
            /clone="CM014g06_r"
            /dev_stage="photoautotrophic growth"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT  131 a 158 c 176 g 94 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 74; Length 583;
Best Local Similarity 90.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 ttgcgcgtacctgacttagcc 20
      ||||| 1 |||||
Db  97 TTGCGGTAGCAGACTTAGCC 78

RESULT 5
AW448957/c
LOCUS      323 bp      mRNA      EST      17-FEB-2000
DEFINITION UI-H-BT3-ake-a-01-0-UI-s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone
            IMAGE:2733960 3', mRNA sequence.
ACCESSION  AW448957
VERSION     AW448957.1  GI:6989733
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 323)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Jan 6, 2000 this sequence version replaced gi:6675688.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES
            Location/Qualifiers
            1..323
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2733960"
            /clone_lib="NCI_CGAP_Sub5"
            /lab_host="DH10B (Life Technologies)"
            /note="Vector: pT7T3D-pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5
            is a substracted library derived from NCI_CGAP_Sub4. The
            NCI_CGAP_Sub5 library had 3 million recombinants. A
            single-stranded DNA preparation of NCI_CGAP_Sub4 was used
            as a tracer in a subtractive hybridization with a driver
            comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
            3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs
            1322376-1323911, 1456008-1456775, 1500552-1502855);
            NCI_CGAP_Kid5 pool 1 LLAM 3338-3342,3722-3725, 3776-3778
            (IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
            1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
            3851-3854 (IMAGE CloneIDs
            1414920-1417991,1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM
            3164-3167, 3716-3720, 3733-3735 (IMAGE CloneIDs

```

```

/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM014g06_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT  133 a 158 c 185 g 107 t
ORIGIN

```

```

Query Match      84.0%; Score 16.8; DB 74; Length 583;
Best Local Similarity 90.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 ttgcgcgtacctgacttagcc 20
      ||||| 1 |||||
Db  97 TTGCGGTAGCAGACTTAGCC 78

```

```

RESULT 5
AW448957/c
LOCUS      323 bp      mRNA      EST      17-FEB-2000
DEFINITION UI-H-BT3-ake-a-01-0-UI-s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone
            IMAGE:2733960 3', mRNA sequence.
ACCESSION  AW448957
VERSION     AW448957.1  GI:6989733
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 323)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Jan 6, 2000 this sequence version replaced gi:6675688.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES
            Location/Qualifiers
            1..323
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2733960"
            /clone_lib="NCI_CGAP_Sub5"
            /lab_host="DH10B (Life Technologies)"
            /note="Vector: pT7T3D-pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5
            is a substracted library derived from NCI_CGAP_Sub4. The
            NCI_CGAP_Sub5 library had 3 million recombinants. A
            single-stranded DNA preparation of NCI_CGAP_Sub4 was used
            as a tracer in a subtractive hybridization with a driver
            comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
            3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs
            1322376-1323911, 1456008-1456775, 1500552-1502855);
            NCI_CGAP_Kid5 pool 1 LLAM 3338-3342,3722-3725, 3776-3778
            (IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
            1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
            3851-3854 (IMAGE CloneIDs
            1414920-1417991,1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM
            3164-3167, 3716-3720, 3733-3735 (IMAGE CloneIDs

```

1257096-1258631,1469064-1470983, 1475592-1476743);  
 NCI\_CGAP\_Pr22 pool 1 L1AM 2457-2459, 2758-2759, 3062-3068  
 (IMAGE Clonoids 985608-986759,1101192-1101959  
 1217928-1220615); NCI\_CGAP\_Co10 pool 1 L1AM 2644-2653,  
 2871-2872 (IMAGE Clonoids  
 1057416-1061255,1144584-1145351). (10% of the driver  
 population), plus a pool of 3,840 arrayed clones from  
 NCI\_CGAP\_Sub1 (IMAGE Clonoids 2708616-2710535) and  
 NCI\_CGAP\_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of  
 the driver population), plus a pool of 11,136 clones from  
 NCI\_CGAP\_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the  
 driver population), plus a pool of 5,472 clones from  
 NCI\_CGAP\_Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the  
 driver population). Subtraction was performed as  
 previously described [Bonaldo, Lennon & Soares (1996):  
 Normalization and Subtraction: Two Approaches to  
 Facilitate Gene Discovery. Genome Research 6, 791-806.  
 TAG\_LIB=NCI\_CGAP\_Pr22  
 TAG\_TISSUE=prostate  
 TAG\_SEQ=AAGTG"

BASE COUNT 76 a 71 c 93 g 83 t  
 ORIGIN

Query Match 79.0%; Score 15.8; DB 81; Length 323;  
 Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ttgcgtacctgacttagcc 20  
 ||||| ||||| ||||| |||||  
 Db 75 TGCCTTACCTGACTTTGCC 57

RESULT 6  
 AQ0909391/c  
 LOCUS GSSTC010112 Trypanosoma cruzi random genomic library Trypanosoma  
 cruzi genomic clone G50F10 5', genomic survey sequence.  
 ACCESSION AQ0909391  
 VERSION AQ0909391.1 GI:6489721  
 KEYWORDS GSS.  
 SOURCE Trypanosoma cruzi.  
 ORGANISM Trypanosoma cruzi.  
 Eukaryota; Eulkenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma; Schizotrypanum.  
 1 (bases 1 to 401)  
 Sanchéz, D.O.  
 Trypanosoma cruzi random genomic sequences  
 Unpublished (1999)  
 On Sep 10, 1998 this sequence version replaced gi:3556080.  
 Contact: Sanchez D.O.  
 Instituto de Investigaciones Biotecnológicas (Univ. Nac. de Gral  
 San Martín)  
 Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24  
 CP(1650) San Martín, Prov. de BS AS. Argentina  
 Tel: (54-11)4752-0021  
 Fax: (54-11)4752-9639  
 Email: gsanchez@inti.gov.ar  
 Seq primer: T7  
 Class: shotgun.

FEATURES  
 source  
 1..401  
 /organism="Trypanosoma cruzi"  
 /strain="Cl-Brener"  
 /db\_xref="taxon:5693"  
 /clone="G50F10"  
 /clone\_lib="Trypanosome cruzi random genomic library"  
 /cell\_type="epimastigote"  
 /note="Vector: pBS(-) (Stratagene)"

BASE COUNT 98 a 102 c 115 g 81 t 5 others  
 ORIGIN

Query Match 79.0%; Score 15.8; DB 92; Length 401;  
 Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgcgtacctgacttagc 19  
 ||||| ||||| ||||| |||||  
 Db 131 TTGCCGTACCTGCCTTGC 113

RESULT 7  
 C18763  
 LOCUS C18763 Human placenta cDNA (TFujiwara) Homo sapiens cDNA clone  
 DEFINITION GEN-567D02 5', mRNA sequence.  
 ACCESSION C18763  
 VERSION C18763.1 GI:1580365  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 421)  
 Fujiwara, T., Hirano, H., Hishigaki, H., Horie, M., Kawai, A., Kuga, Y.,  
 Kyushiki, H., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y.,  
 Shinomiya, H., Suzuki, M., Takaichi, A., Takeda, S., Watanabe, T.,  
 Maekawa, H., Nakamura, Y. and Takahashi, E.  
 Otsuka cDNA project  
 Unpublished (1996)  
 On Sep 12, 1996 this sequence version replaced gi:1993736.  
 Contact: Tsutomu Fujiwara  
 Otsuka GEN Research Institute  
 Otsuka Pharmaceutical Co., Ltd  
 463-10 kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan  
 Tel: 0886-65-2888  
 Fax: 0886-37-1035.

FEATURES  
 Location/Qualifiers  
 source  
 1..421  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="GEN-567D02"  
 /clone\_lib="Human placenta cDNA (TFujiwara)"  
 /tissue\_type="placenta"

BASE COUNT 115 a 100 c 116 g 90 t  
 ORIGIN

Query Match 79.0%; Score 15.8; DB 28; Length 421;  
 Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ttgcgtacctgacttagcc 20  
 ||||| ||||| ||||| |||||  
 Db 359 TGCCGTTCTGCCTTAGCC 377

RESULT 8  
 AW330086/c  
 LOCUS TENU4832 T.cruzi epimastigote normalized cDNA Library Trypanosoma  
 cruzi cDNA clone 43e14 5', mRNA sequence.  
 ACCESSION AW330086  
 VERSION AW330086.1 GI:6807144  
 KEYWORDS EST.  
 SOURCE Trypanosoma cruzi.  
 ORGANISM Trypanosoma cruzi.  
 Eukaryota; Eulkenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma; Schizotrypanum.  
 1 (bases 1 to 574)  
 Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M.,  
 Urmenyi, T.P., Rondinelli, E., Pettersson, U., Andersson, B. and  
 Aslund, L.  
 Gene survey of the pathogenic protozoan Trypanosoma cruzi  
 Unpublished (2000)

COMMENT On May 18, 1998 this sequence version replaced gi:3136933.  
 Contact: Aslund L  
 Department of Medical Genetics  
 Uppsala University  
 Biomedical Center, Box 589, S-751 23 Uppsala, Sweden  
 Tel: 46 18 471 45 85  
 Fax: 46 18 52 68 49  
 Email: lena.aslund@medgen.uu.se  
 Seq primer: T7 primer  
 High quality sequence stop: 574.  
 Location/Qualifiers  
 1. .574  
 /organism="Trypanosoma cruzi"  
 /strain="Cl-Brenner"  
 /db\_xref="taxon:5693"  
 /clone="43e14"  
 /clone\_lib="T. cruzi epimastigote normalized cDNA Library"  
 /cell\_type="epimastigote"  
 /note="cDNA library constructed with oligo dt primed  
 epimastigote mRNA and cloned in pT7t318D phagemid with  
 modified polylinker (Pharmacia)"  
 129 a 135 c 172 g 138 t

BASE COUNT  
 ORIGIN

Query Match 79.0%; Score 15.8; DB 80; Length 574;  
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgcgcgtacctgactgac 19  
 ||||| ||||| ||||| |||||  
 Db 258 TTGCGGTACTGCTGCGTGC 240

RESULT 9  
 AA911667  
 LOCUS  
 DEFINITION O149f04.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 IMAGE:1526815 3', mRNA sequence.  
 AA911667  
 VERSION  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 579)  
 NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Jan 19, 1998 this sequence version replaced gi:2285695.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL : contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 662 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 194.  
 Location/Qualifiers  
 1. .579  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1526815"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT7t3D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NC1-CGAP-GCB1) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo.  
 86 a 196 c 201 g 96 t

BASE COUNT  
 ORIGIN

Query Match 79.0%; Score 15.8; DB 40; Length 579;  
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgcgcgtacctgactgac 19  
 ||||| ||||| ||||| |||||  
 Db 521 TTGCGGTCTCTGCTTACG 539

RESULT 10  
 AA929849  
 LOCUS  
 DEFINITION vs55c09.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
 IMAGE:1150192 5', similar to gb:X54199 PHOSPHORIBOSYLAMINE--GLYCINE  
 LIASE (HUMAN); gb:U01024 Mus musculus glycineamide ribonucleotide  
 synthetase (MOUSE);, mRNA sequence.  
 AA929849  
 VERSION  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM  
 Mus musculus  
 house mouse.  
 1 (bases 1 to 599)  
 REFERENCE  
 AUTHORS  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilsson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 On Jan 19, 1998 this sequence version replaced gi:2285441.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of MedicineP  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL : contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:623400  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 63.  
 Location/Qualifiers  
 1. .599  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1150192"  
 /clone\_lib="Stratagene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skin; Vector: pBluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. Whole skin from 11 week old C57BL/6 female mice.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGCACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"  
 146 a 154 c 150 g 148 t

BASE COUNT  
 ORIGIN

```

Query Match      79.0%; Score 15.8; DB 40; Length 599;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgcgtacctgacttagcc 20
|||||
Db 205 TGGCGTACCTGGCTTAGAC 223

RESULT 11
AQ473994/c
LOCUS      605 bp      DNA      GSS      23-APR-1999
DEFINITION CITBI-E1-2583N8.TF CITBI-E1 Homo sapiens genomic clone 2583N8,
genomic survey sequence.
ACCESSION  AQ473994
VERSION     AQ473994.1 GI:4652255
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 605)
AUTHORS   Zhao,S., Adams M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE     Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
            Map Building
JOURNAL   Unpublished (1997)
COMMENT   Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
            Location/Qualifiers
                1..605
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="2583N8"
                /clone_lib="CITBI-E1"
                /sex="male"
                /cell_type="sperm"
                /note="Vector: pBelBAC11; Site_1: EcoRI; Site_2: EcoRI;
            Caltech Human BAC Library D"
BASE COUNT 171 a 138 c 136 g 160 t
ORIGIN

Query Match      77.0%; Score 15.4; DB 107; Length 605;
Best Local Similarity 94.1%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ttgcgtacctgacttagc 17
|||||
Db 308 TTGCAGTACCTGACTTA 292

RESULT 12
AQ528466/c
LOCUS      710 bp      DNA      GSS      18-MAY-1999
DEFINITION RPCI-11-349N7.TV RPCI-11 Homo sapiens genomic clone RPCI-11-349N7,
genomic survey sequence.
ACCESSION  AQ528466
VERSION     AQ528466.1 GI:4840579
KEYWORDS   GSS.
SOURCE     human.

```

```

ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 710)
AUTHORS   Zhao,S., Adams M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
JOURNAL   Unpublished (1997)
COMMENT   On Sep 10, 1998 this sequence version replaced gi:3554810.
            Other_GSSs: RPCI-11-349N7.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
            Location/Qualifiers
                1..710
                /organism="Homo sapiens"
                /db_xref="GDB:7633950"
                /db_xref="taxon:9606"
                /clone="RPCI-11-349N7"
                /clone_lib="RPCI-11"
                /sex="Male"
                /cell_type="Lymphocytes"
                /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
            RPCI11 Human Male BAC Library"
BASE COUNT 189 a 173 c 157 g 191 t
ORIGIN

Query Match      77.0%; Score 15.4; DB 108; Length 710;
Best Local Similarity 94.1%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ttgcgtacctgacttagc 17
|||||
Db 303 TTGCAGTACCTGACTTA 287

RESULT 13
AA957921/c
LOCUS      166 bp      mRNA      EST      04-JUL-1999
DEFINITION UI-R-E1-fv-c-09-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone
UI-R-E1-fv-c-09-0-UI 3', mRNA sequence.
ACCESSION  AA957921
VERSION     AA957921.1 GI:4277811
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE  1 (bases 1 to 166)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704477
COMMENT   On May 7, 1998 this sequence version replaced gi:3121616.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA

```

Tel: 319 335 8250  
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 12-day-Embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1780092

Seq primer: M13 Forward

POLYA-No.

Location/Qualifiers

1. .166

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-E1-fv-c-09-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-E1 library is a subtracted library derived from the UI-R-E0 library. The UI-R-E0 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-E1) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-E0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-E0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-E1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

40 a 28 c 49 g 49 t

BASE COUNT  
ORIGIN

Query Match 76.0%; Score 15.2; DB 46; Length 166;

Best Local Similarity 85.0%; Pred. NO. 2.9e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttgccgtacctgacttagcc 20

||||| ||||| ||||| |||||

Db 134 TTGCCCTACCTGACCTAGAC 115

RESULT 14

B71840/c

LOCUS B71840 183 bp DNA GSS 08-APR-1999

DEFINITION RPC111-11C6.TP RPCI-11 Homo sapiens genomic clone RPCI-11-11C6, genomic survey sequence.

ACCESSION B71840

VERSION B71840.1 GI:2711061

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 183)

AUTHORS Adams M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K.,

Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and

Venter, J.C.

Use of BAC End Sequences for Sequence-Ready Map Building

Unpublished (1997)

COMMENT

Other\_GSSs: RPC111-11C6.TV

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html

Seq primer: SP6

Class: BAC ends.

FEATURES  
source

Location/Qualifiers

1. .183

/organism="Homo sapiens"

/db\_xref="GDB:7503893"

/db\_xref="taxon:9606"

/clone="RPCI-11-11C6"

/clone\_lib="RPCI-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPC111 Human Male BAC Library"

62 a 44 c 44 g 33 t

BASE COUNT  
ORIGIN

Query Match 76.0%; Score 15.2; DB 84; Length 183;

Best Local Similarity 85.0%; Pred. No. 3e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ttgccgtacctgacttagcc 20

||||| ||||| ||||| |||||

Db 21 TTGCCCTTCATGACTTAGCC 2

RESULT 15

AI045089/c

LOCUS

AI045089 243 bp mRNA EST 11-FEB-1999

DEFINITION UI-R-C1-kf-f-12-0-UI-s1 UI-R-C1 Rattus norvegicus cDNA clone

UI-R-C1-kf-f-12-0-UI 3', mRNA sequence.

ACCESSION AI045089

VERSION AI045089.1 GI:3291908

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 243)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 9704477

COMMENT On Jan 9, 1998 this sequence version replaced gi:937497.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 12-day-Embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics

Seq primer: M13 Forward.

Location/Qualifiers

FEATURES

source

```
1. .243
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C1-Kf-f-12-0-UI"
/clone_lib="UI-R-C1"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The UI-R-C1
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The UI-R-E1 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C1) was constructed as follows: PCR
amplified cDNA inserts from UI-R-C0 clones from which 3'
ESTs had been derived was used as a driver in a
hybridization with the UI-R-C0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-C1
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."
51 a 53 c 78 g 61 t
```

BASE COUNT  
ORIGIN

Query Match 76.0%; Score 15.2; DB 41; Length 243;  
Best Local Similarity 85.0%; Pred. No. 3.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20  
|||||  
Db 134 TTGCCCTACTGACCTAGAC 115

Search completed: May 23, 2000, 09:21:43  
Job time: 7770 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:40:21 ; Search time 71.7 Seconds  
(without alignments)

36.258 Million cell updates/sec

Title: US-08-945-805-2

Perfect score: 20

Sequence: 1 ttgccgtacctgacttagcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	79.0	39	2	US-08-484-192-23
2	14.2	71.0	1394	5	US-08-880-499-1
3	14.2	71.0	1394	5	US-08-880-499-2
4	14	70.0	223	5	US-08-735-545-23
5	13.8	69.0	1482	1	US-08-341-916-1
6	13.8	69.0	1482	2	US-08-805-166-1
7	13.8	69.0	1482	2	US-08-805-169-1
8	13.8	69.0	1482	3	US-08-957-365-1
9	13.8	69.0	1482	4	US-08-957-364-1
10	13.8	69.0	6605	2	US-08-769-309A-4
11	13.8	69.0	43280	3	US-08-804-227C-1
12	13.6	68.0	238	1	US-07-914-284A-1
13	13.6	68.0	238	6	PCT-US93-06645-1
14	13.6	68.0	958	3	US-08-632-514C-9
15	13.6	68.0	973	2	US-08-532-309-1
16	13.6	68.0	1689	1	US-08-442-884-2
17	13.6	68.0	2614	3	US-08-795-868-15
18	13.6	68.0	6363	3	US-08-929-967-6
19	13.6	68.0	8323	1	US-08-110-300A-8
20	13.6	68.0	8323	4	US-08-886-642-8
21	13.6	68.0	8323	6	PCT-US93-08041-8
22	13.6	68.0	10367	1	US-08-110-300A-9
23	13.6	68.0	10367	4	US-08-886-642-9
24	13.6	68.0	10367	6	PCT-US93-08041-9
25	13.2	66.0	527	5	US-08-714-918-26
26	13.2	66.0	1453	3	US-08-169-948B-9
27	13.2	66.0	1453	3	US-08-448-873-9

c	28	13.2	66.0	1778	6	PCT-US91-02958-9	Sequence 9, Appl
	29	13.2	66.0	1820	4	US-08-389-564B-17	Sequence 17, Appl
	30	13.2	66.0	1820	5	US-08-466-047B-17	Sequence 17, Appl
	31	13.2	66.0	2018	5	US-08-714-918-16	Sequence 16, Appl
c	32	13.2	66.0	2219	1	US-08-290-979A-7	Sequence 7, Appl
	33	13.2	66.0	2385	1	US-08-393-333-1	Sequence 1, Appl
	34	13.2	66.0	3186	5	US-08-863-102-3	Sequence 3, Appl
c	35	13.2	66.0	3213	3	US-08-633-770A-4	Sequence 4, Appl
	36	13.2	66.0	4248	5	US-08-678-614-1	Sequence 1, Appl
c	37	13.2	66.0	4670	3	US-08-633-770A-12	Sequence 12, Appl
	38	13.2	66.0	6678	5	US-08-816-617A-1	Sequence 1, Appl
	39	13.2	66.0	8051	4	US-08-576-626A-2	Sequence 2, Appl
	40	13	65.0	1950	5	US-08-685-466C-1	Sequence 1, Appl
c	41	13	65.0	2526	4	US-08-702-572-15	Sequence 15, Appl
	42	12.8	64.0	1080	1	US-08-137-627-3	Sequence 3, Appl
	43	12.8	64.0	1080	3	US-08-865-348-3	Sequence 3, Appl
c	44	12.8	64.0	1135	2	US-08-624-125-12	Sequence 12, Appl
c	45	12.8	64.0	1165	2	US-08-624-125-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-08-484-192-23  
; Sequence 23, Application US/08484192  
; Patent No. 5756291  
; GENERAL INFORMATION:  
; APPLICANT: GRIFFIN, LINDA C.  
; APPLICANT: ALBRECHT, GLENN  
; APPLICANT: LATHAM, JOHN  
; APPLICANT: LEUNG, LAWRENCE  
; APPLICANT: VERMAAS, ERIC  
; APPLICANT: TOOLE, JOHN J.  
; TITLE OF INVENTION: APTAMERS SPECIFIC FOR BIOMOLECULES AND  
; NUMBER OF SEQUENCES: 181  
; TITLE OF INVENTION: METHODS OF MAKING  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,192  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/934,387  
; FILING DATE: 21-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GRACEY, NANCY J.  
; REGISTRATION NUMBER: 28,216  
; REFERENCE/DOCKET NUMBER: 246102002221  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-813-5600  
; TELEFAX: 415-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: misc\_difference  
; LOCATION: replace(20, "")

; OTHER INFORMATION: /note: "This is a 60 nucleotide  
; OTHER INFORMATION: stretch of random sequences."  
US-08-484-192-23

Query Match 79.0%; Score 15.8; DB 2; Length 39;  
Best Local Similarity 85.0%; Pred. No. 6.4;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20  
||||| ||||| |||  
Db 14 TTGCGGNACCTGAATTCGCC 33

RESULT 2  
US-08-880-499-1/c  
; Sequence 1, Application US/08880499  
; Patent No. 6037523  
; GENERAL INFORMATION:  
; APPLICANT: Albertson, Marc C.  
; APPLICANT: Fox, Tim W.  
; APPLICANT: Carl, Garnaat W.  
; APPLICANT: Huffman, Gary A.  
; APPLICANT: Kendall, Timmy L.  
; TITLE OF INVENTION: MALE TISSUE-PREFERRED REGULATORY REGION  
; TITLE OF INVENTION: AND METHOD OF USING SAME  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.  
; STREET: Darwin Building, 7100 N.W. 62nd Ave., P.O.  
; STREET: Box 1000  
; CITY: Johnston  
; STATE: Iowa  
; COUNTRY: USA  
; ZIP: 50131  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/880,499  
; FILING DATE: CONCURRENTLY HERewith  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweeney, Patricia A.  
; REGISTRATION NUMBER: 32,733  
; REFERENCE/DOCKET NUMBER: 0578  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (515) 248-4800  
; TELEFAX: (515) 248-4844  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1394 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-880-499-1

Query Match 71.0%; Score 14.2; DB 5; Length 1394;  
Best Local Similarity 84.2%; Pred. No. 59;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttagcc 20  
||||| ||||| |||  
Db 1149 TGCCGAACCTGACCTAGGC 1131

RESULT 3  
US-08-880-499-2/c  
; Sequence 2, Application US/08880499

; Patent No. 6037523  
; GENERAL INFORMATION:  
; APPLICANT: Albertson, Marc C.  
; APPLICANT: Fox, Tim W.  
; APPLICANT: Carl, Garnaat W.  
; APPLICANT: Huffman, Gary A.  
; APPLICANT: Kendall, Timmy L.  
; TITLE OF INVENTION: MALE TISSUE-PREFERRED REGULATORY REGION  
; TITLE OF INVENTION: AND METHOD OF USING SAME  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.  
; STREET: Darwin Building, 7100 N.W. 62nd Ave., P.O.  
; STREET: Box 1000  
; CITY: Johnston  
; STATE: Iowa  
; COUNTRY: USA  
; ZIP: 50131  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/880,499  
; FILING DATE: CONCURRENTLY HERewith  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweeney, Patricia A.  
; REGISTRATION NUMBER: 32,733  
; REFERENCE/DOCKET NUMBER: 0578  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (515) 248-4800  
; TELEFAX: (515) 248-4844  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1394 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-880-499-2

Query Match 71.0%; Score 14.2; DB 5; Length 1394;  
Best Local Similarity 84.2%; Pred. No. 59;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttagcc 20  
||||| ||||| |||  
Db 1149 TGCCGAACCTGACCTAGGC 1131

RESULT 4  
US-08-735-545-23  
; Sequence 23, Application US/08735545  
; Patent No. 6025131  
; GENERAL INFORMATION:  
; APPLICANT: Van Dyk, Tina K.  
; APPLICANT: LaRossa, Robert Alan  
; TITLE OF INVENTION: A Facile Method for  
; TITLE OF INVENTION: Identifying Regulated  
; TITLE OF INVENTION: Promoters  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS  
; ADDRESSEE: AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1  
; SOFTWARE: MICROSOFT WORD 2.0C  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735,545  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FLOYD, LINDA A.  
; REGISTRATION NUMBER: 33,692  
; REFERENCE/DOCKET NUMBER: CR-9989  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-892-8112  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; STRAIN: dpd3509 upper  
; US-08-735-545-23

Query Match 70.0%; Score 14; DB 5; Length 223;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 gtacctgacttagc 19  
|||||  
Db 98 GTACCTGACTTAGC 111

RESULT 5  
; US-08-341-916-1  
; Sequence 1, Application US/08341916  
; Patent No. 5614609  
; GENERAL INFORMATION:  
; APPLICANT: Ib ez, Carlos F.  
; APPLICANT: Ryd n, Mikael  
; APPLICANT: J rnvall, Henrik  
; TITLE OF INVENTION: A No. 5614609el Serine Threonine Kinase Receptor  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/341,916  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldstein, Jorge A.  
; REGISTRATION NUMBER: 29,021  
; REFERENCE/DOCKET NUMBER: 1459.0230001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)371-2600  
; TELEFAX: (202)371-2540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1482 base pairs  
; TYPE: nucleic acid

; STRANDEDNESS: both  
; TOPOLOGY: both  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1482  
; US-08-341-916-1

Query Match 69.0%; Score 13.8; DB 1; Length 1482;  
Best Local Similarity 88.2%; Pred. No. 95;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcctgactgacttag 18  
|||||  
Db 1011 TGCCATAGCTGACTTAG 1027

RESULT 6  
; US-08-805-166-1  
; Sequence 1, Application US/08805166  
; Patent No. 5789565  
; GENERAL INFORMATION:  
; APPLICANT: Ib ez, Carlos F.  
; APPLICANT: Ryd n, Mikael  
; APPLICANT: J rnvall, Henrik  
; TITLE OF INVENTION: A No. 5789565el Serine Threonine Kinase Receptor  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/805,166  
; FILING DATE: 24-FEB-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/341,916  
; FILING DATE: 15-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldstein, Jorge A.  
; REGISTRATION NUMBER: 29,021  
; REFERENCE/DOCKET NUMBER: 1459.0230001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)371-2600  
; TELEFAX: (202)371-2540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1482 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1482  
; US-08-805-166-1

Query Match 69.0%; Score 13.8; DB 2; Length 1482;  
Best Local Similarity 88.2%; Pred. No. 95;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcctgactgacttag 18  
|||||  
Db 1011 TGCCATAGCTGACTTAG 1027

```
RESULT 7
US-08-805-169-1
; Sequence 1, Application US/08805169
; Patent No. 5811245
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J Invall, Henrik
; TITLE OF INVENTION: A No. 5811245el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,169
; FILING DATE: 24-FEB-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/341,916
; FILING DATE: 15-NOV-1994
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1459.0230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1482 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1482
; US-08-805-169-1

Query Match 69.0%; Score 13.8; DB 2; Length 1482;
Best Local Similarity 88.2%; Pred. No. 95;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttag 18
||||| |||||
Db 1011 TGGCATAGCTGACTTAG 1027

RESULT 8
US-08-957-365-1
; Sequence 1, Application US/08957365
; Patent No. 5891638
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J Invall, Henrik
; TITLE OF INVENTION: A No. 5891638el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,364
; FILING DATE: 24-FEB-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/805,169
; FILING DATE: 15-NOV-1994
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1459.0230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1482 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1482
; US-08-957-365-1

Query Match 69.0%; Score 13.8; DB 2; Length 1482;
Best Local Similarity 88.2%; Pred. No. 95;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttag 18
||||| |||||
Db 1011 TGGCATAGCTGACTTAG 1027

RESULT 9
US-08-957-364-1
; Sequence 1, Application US/08957364
; Patent No. 5976815
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J Invall, Henrik
; TITLE OF INVENTION: A No. 5976815el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,364
; FILING DATE: 24-FEB-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/805,169
; FILING DATE: 15-NOV-1994
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1459.0230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1482 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1482
; US-08-957-365-1
```

```
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,365
; FILING DATE: 24-OCT-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/805,166
; FILING DATE: 24-FEB-1997
; APPLICATION NUMBER: 08/341,916
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1459.0230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1482 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1482
; US-08-957-365-1

Query Match 69.0%; Score 13.8; DB 3; Length 1482;
Best Local Similarity 88.2%; Pred. No. 95;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgccgtacctgacttag 18
||||| |||||
Db 1011 TGGCATAGCTGACTTAG 1027

RESULT 9
US-08-957-364-1
; Sequence 1, Application US/08957364
; Patent No. 5976815
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J Invall, Henrik
; TITLE OF INVENTION: A No. 5976815el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,364
; FILING DATE: 24-FEB-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/805,169
; FILING DATE: 24-FEB-1997
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/341,916
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1459.0230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1482 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1482
; US-08-957-364-1

Query Match 69.0%; Score 13.8; DB 4; Length 1482;
Best Local Similarity 88.2%; Pred. NO. 95;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcgcgtacctgacttag 18
||||| || ||||| |||||
Db 1011 TGCCATAGCTGACTTAG 1027

RESULT 10
US-08-769-309A-4
; Sequence 4, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:

Query Match 69.0%; Score 13.8; DB 3; Length 43280;

; NAME/KEY: CDS
; LOCATION: 192..5531
; US-08-769-309A-4

Query Match 69.0%; Score 13.8; DB 2; Length 6605;
Best Local Similarity 88.2%; Pred. NO. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcgcgtacctgacttag 18
||||| ||||| ||||| |||
Db 3269 TGGCGTACCTGACATAG 3285

RESULT 11
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
; US-08-804-227C-1

Query Match 69.0%; Score 13.8; DB 3; Length 43280;
```

Best Local Similarity 88.2%; Pred. No. 1.2e+02; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 2;

QY 4 cccgtacctgacttagcc 20  
||| ||||| |||||

DB 353 CCGGACCTGCATAGCC 369

## RESULT 12

US-07-914-284A-1  
; Sequence 1, Application US/07914284A  
; Patent No. 5489524  
; GENERAL INFORMATION:  
; APPLICANT: Chow, Terry Y.-K.  
; APPLICANT: Resnick, Michael A.  
; APPLICANT: Perkins, Edward  
; TITLE OF INVENTION: A CHIMERIC PROTEIN THAT HAS A HUMAN RHO  
; MOTIF AND DEOXYRIBONUCLEASE ACTIVITY  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/914,284A  
; FILING DATE: 14-JUL-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/674,801  
; FILING DATE: 26-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E.  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: NIH022.022CP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 238 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; CLONE: HuNUC-T7  
; PCT-US93-06645-1

Query Match 68.0%; Score 13.6; DB 6; Length 238;  
Best Local Similarity 80.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20  
||| ||||| ||| ||| |||  
DB 50 TTGCCGTTGCTGCTTTGCC 69

## RESULT 14

US-08-632-514C-9  
; Sequence 9, Application US/08632514C  
; Patent No. 5834234  
; GENERAL INFORMATION:  
; APPLICANT: GALLO, Gregory J.  
; TITLE OF INVENTION: APOPTOSIS ASSOCIATED PROTEIN BBK  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hale and Dorr  
; STREET: 1455 Pennsylvania Avenue, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/632,514C  
; FILING DATE: 29-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WIXON, Henry N.  
; REGISTRATION NUMBER: 32,073  
; REFERENCE/DOCKET NUMBER: 104322.188  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)942-8459  
; TELEFAX: (202)942-8484

Query Match 68.0%; Score 13.6; DB 1; Length 238;  
Best Local Similarity 80.0%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ttgccgtacctgacttagcc 20  
||| ||||| ||| ||| |||

DB 50 TTGCCGTTGCTGCTTTGCC 69

## RESULT 13

PCT-US93-06645-1  
; Sequence 1, Application PC/TUS9306645  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: A CHIMERIC PROTEIN THAT  
; MOTIF AND

; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 958 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-632-514C-9

Query Match 68.0%; Score 13.6; DB 3; Length 958;  
Best Local Similarity 80.0%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttgccgtacctgacttagcc 20  
||||| ||||| |||||  
Db 902 TTGCCCTACCTGTTTGTGCC 921

RESULT 15

US-08-532-309-1  
; Sequence 1, Application US/08532309  
; Patent No. 5795975

; GENERAL INFORMATION:  
; APPLICANT: WALLACH, David  
; APPLICANT: KEMPER, Oliver  
; APPLICANT: KUHNERT, Peter  
; TITLE OF INVENTION: TNF RECEPTOR PROMOTER  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States of America  
; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/532.309  
; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: BROWDY, Roger L.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: WALLACH-11A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 628-5197

; TELEFAX: (202) 737-3528

; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 973 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-532-309-1

Query Match 68.0%; Score 13.6; DB 2; Length 973;  
Best Local Similarity 80.0%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ttgccgtacctgacttagcc 20  
||||| ||||| |||||  
Db 479 TTCCGGATCTGACTTAACC 498

Search completed: May 23, 2000, 09:40:27  
Job time: 8766 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:40:01 ; Search time 693.2 Seconds  
(without alignments)  
-28.067 Million cell updates/sec

Title: US-08-945-805-3  
Perfect score: 20  
Sequence: 1 ggaacttcctcctaaaggagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pl1:\*
- 8: gb\_pl2:\*
- 9: gb\_pr1:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: gb\_ro:\*
- 13: gb\_sts:\*
- 14: gb\_sy:\*
- 15: gb\_un:\*
- 16: gb\_vi:\*
- 17: em\_fun:\*
- 18: em\_humi:\*
- 19: em\_hum2:\*
- 20: em\_in:\*
- 21: em\_om:\*
- 22: em\_or:\*
- 23: em\_ov:\*
- 24: em\_pat:\*
- 25: em\_ph:\*
- 26: em\_pl:\*
- 27: em\_ro:\*
- 28: em\_sts:\*
- 29: em\_sy:\*
- 30: em\_un:\*
- 31: em\_vi:\*
- 32: gb\_htg1:\*
- 33: gb\_htg2:\*
- 34: gb\_inl:\*
- 35: gb\_in2:\*
- 36: em\_ba1:\*
- 37: em\_ba2:\*
- 38: em\_hum3:\*
- 39: em\_hum4:\*
- 40: gb\_pr4:\*
- 41: gb\_htg3:\*
- 42: gb\_htg4:\*
- 43: gb\_htg5:\*
- 44: gb\_htg6:\*

- 45: gb\_htg7:\*
- 46: em\_htg1:\*
- 47: em\_htg2:\*
- 48: em\_htg3:\*
- 49: em\_hum5:\*
- 50: gb\_pl3:\*
- 51: gb\_pr5:\*
- 52: gb\_htg8:\*
- 53: gb\_htg9:\*
- 54: gb\_htg10:\*
- 55: gb\_htg11:\*
- 56: gb\_htg12:\*
- 57: gb\_htg13:\*
- 58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	19	95.0	340860	56	AC008753	AC008753 Homo sapi
C 2	18	90.0	87180	7	AB023044	AB023044 Arabidops
C 3	17.4	87.0	127339	12	AC000096	AC000096 Mus muscu
4	17.4	87.0	178964	43	AC012297	AC012297 Mus muscu
5	17.4	87.0	196435	53	AC019079	AC019079 Homo sapi
C 6	17	85.0	64714	7	AP000732	AP000732 Arabidops
C 7	17	85.0	157644	55	AC023483	AC023483 Homo sapi
C 8	17	85.0	165958	43	AC016092	AC016092 Homo sapi
C 9	17	85.0	176475	55	AC022800	AC022800 Homo sapi
C 10	17	85.0	202478	40	AC006560	AC006560 Homo sapi
C 11	16.8	84.0	39436	9	AB000879	AB000879 Homo sapi
C 12	16.8	84.0	44710	40	AC004184	AC004184 Homo sapi
C 13	16.8	84.0	64406	41	AC011827	AC011827 Homo sapi
C 14	16.8	84.0	100000	9	AP000506	AP000506 Homo sapi
C 15	16.8	84.0	135953	44	AC018412	AC018412 Homo sapi
C 16	16.8	84.0	148960	9	AB000882	AB000882 Homo sapi
17	16.8	84.0	186666	43	AC012564	AC012564 Homo sapi
18	16.8	84.0	186872	33	AL133350	AL133350 Homo sapi
19	16.8	84.0	291131	33	HSAL175J10	AL121927 Homo sapi
C 20	16.4	82.0	80470	42	AC015678	AC015678 Homo sapi
21	16.4	82.0	107978	42	AC012129	AC012129 Homo sapi
22	16.4	82.0	136945	45	AC021974	AC021974 Homo sapi
C 23	16.4	82.0	145072	41	AC009758	AC009758 Homo sapi
24	16.4	82.0	146248	42	AC010930	AC010930 Homo sapi
25	16.4	82.0	148647	32	AL137001	AL137001 Homo sapi
C 26	16.4	82.0	149869	40	AC006026	AC006026 Homo sapi
27	16.4	82.0	164236	41	AC009612	AC009612 Homo sapi
28	16.4	82.0	166826	43	AC012580	AC012580 Homo sapi
C 29	16.4	82.0	187040	52	AC021554	AC021554 Homo sapi
30	16.4	82.0	188567	41	AC008176	AC008176 Homo sapi
C 31	16.4	82.0	191549	51	AC010582	AC010582 Homo sapi
32	16.4	82.0	217419	32	CNS01D0V	AL136040 Homo sapi
33	16.4	82.0	233866	33	AL133350	AL133350 Homo sapi
34	16.4	82.0	267156	40	U66059	U66059 Human germ
35	16.4	82.0	348194	33	HSJ1164C1	AL121781 Homo sapi
36	16	80.0	1188	4	XLXWNT5C	X73510 X.laervis Xv
37	16	80.0	6915	1	SMA6085	AJ006085 Shewanell
C 38	16	80.0	41624	11	AC005680	AC005680 Homo sapi
C 39	16	80.0	46885	44	AC021182	AC021182 Homo sapi
C 40	16	80.0	50502	11	HSAA494O16	AL117328 Human DNA
C 41	16	80.0	100427	11	AC004048	AC004048 Homo sapi
C 42	16	80.0	188037	56	AC023599	AC023599 Homo sapi
43	15.8	79.0	14443	2	AE001742	AE001742 Thermotoq
C 44	15.8	79.0	32163	45	AC008828	AC008828 Homo sapi
C 45	15.8	79.0	59348	42	AC013655	AC013655 Homo sapi

ALIGNMENTS

[illegible]

\* 61702 63244: contig of 1543 bp in length  
\* gap of unknown length  
\* 63245 64423: contig of 1179 bp in length  
\* gap of unknown length  
\* 64424 65459: contig of 1036 bp in length  
\* gap of unknown length  
\* 65460 66467: contig of 1008 bp in length  
\* gap of unknown length  
\* 66468 67873: contig of 1406 bp in length  
\* gap of unknown length  
\* 67874 69177: contig of 1304 bp in length  
\* gap of unknown length  
\* 69178 70239: contig of 1062 bp in length  
\* gap of unknown length  
\* 70240 71785: contig of 1546 bp in length  
\* gap of unknown length  
\* 71786 72971: contig of 1186 bp in length  
\* gap of unknown length  
\* 72972 74559: contig of 1588 bp in length  
\* gap of unknown length  
\* 74560 76097: contig of 1538 bp in length  
\* gap of unknown length  
\* 76098 77414: contig of 1317 bp in length  
\* gap of unknown length  
\* 77415 78545: contig of 1131 bp in length  
\* gap of unknown length  
\* 78546 79547: contig of 1002 bp in length  
\* gap of unknown length  
\* 79548 80660: contig of 1113 bp in length  
\* gap of unknown length  
\* 80661 82228: contig of 1568 bp in length  
\* gap of unknown length  
\* 82229 83263: contig of 1035 bp in length  
\* gap of unknown length  
\* 83264 84609: contig of 1346 bp in length  
\* gap of unknown length  
\* 84610 85883: contig of 1274 bp in length  
\* gap of unknown length  
\* 85884 87576: contig of 1693 bp in length  
\* gap of unknown length  
\* 87577 89044: contig of 1468 bp in length  
\* gap of unknown length  
\* 89045 90373: contig of 1329 bp in length  
\* gap of unknown length  
\* 90374 91599: contig of 1226 bp in length  
\* gap of unknown length  
\* 91600 93025: contig of 1426 bp in length  
\* gap of unknown length  
\* 93026 94443: contig of 1418 bp in length  
\* gap of unknown length  
\* 94444 95482: contig of 1039 bp in length  
\* gap of unknown length  
\* 95483 97609: contig of 2127 bp in length  
\* gap of unknown length  
\* 97610 99401: contig of 1792 bp in length  
\* gap of unknown length  
\* 99402 100579: contig of 1178 bp in length

Query Match 95.0%; Score 19; DB 56; Length 340860;  
Best Local Similarity 100.0%; Pred. No. 2.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 gaacttcctaaaggagg 20  
|||||  
Db 125194 GAACCTCCCTAAGGAGG 125176

RESULT 2  
AB023044/c 87180 bp DNA PLN 20-NOV-1999  
LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone: MWD22,  
DEFINITION complete sequence.  
ACCESSION AB023044

VERSION AB023044.1 GI:4220643  
KEYWORDS HTG.  
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui P1  
clone:MWD22.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis.  
REFERENCE 1 (sites)  
AUTHORS Nakamura,Y.  
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. XI  
JOURNAL Unpublished (1998)  
REFERENCE 2 (bases 1 to 87180)  
AUTHORS Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (01-FEB-1999) to the DDBJ/EMBL/GenBank databases.  
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of  
Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan  
(E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,  
Fax:+81-438-52-3934)  
FEATURES Location/Qualifiers  
source 1..87180  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="MWD22"  
/clone\_lib="Mitsui P1"  
BASE COUNT 26079 a 16236 c 16514 g 28351 t  
ORIGIN  
  
Query Match 90.0%; Score 18; DB 7; Length 87180;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Oy 2 gaacttcctaaaggagg 19  
|||||  
Db 27820 GAACCTCCCTAAGGAGG 27803  
  
RESULT 3  
AC000096  
LOCUS AC000096 127339 bp DNA ROD 15-FEB-2000  
DEFINITION Mus musculus Chromosome 16 BAC Clone bd3-6 Syntenic To Homo sapiens  
22q11.2 DSCR Region, complete sequence.  
ACCESSION AC000096  
VERSION AC000096.13 GI:6119551  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 127339)  
AUTHORS Galili,N., Baldwin,S., Lund,J., Reeves,R., Gong,W., Wang,Z.,  
Roe,B.A., Emanuel,B.S., Nayak,S., Mickanin,C., Budarf,M.L. and  
Buck,C.A.  
TITLE A Region of Mouse Chromosome 16 is Syntenic to the DiGeorge,  
Velo-Cardio-Facial Syndrome Minimal Critical Region  
JOURNAL Unpublished  
REMARK The genes were identified by comparing with human genomic and cDNA  
sequences and RT-PCR of 12 day post conception mouse embryos total  
RNA  
REFERENCE 2 (bases 1 to 127339)  
AUTHORS Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-1996) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 3 (bases 1 to 127339)  
AUTHORS Roe,B.A.  
TITLE Direct Submission

```

JOURNAL      Submitted (31-JUL-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE     4 (bases 1 to 127339)
AUTHORS      Roe,B.A.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE     5 (bases 1 to 127339)
AUTHORS      Roe,B.A.
TITLE        Direct Submission
JOURNAL      Submitted (26-OCT-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE     6 (bases 1 to 127339)
AUTHORS      Roe,B.A.
TITLE        Direct Submission
JOURNAL      Submitted (15-FEB-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
COMMENT      On Oct 26, 1999 this sequence version replaced gi:5579284.
FEATURES     Location/Qualifiers
              source
                1..127339
                  /organism="Mus musculus"
                  /strain="129/Sv EV cell line CJ7"
                  /db_xref="taxon:10090"
                  /chromosome="16"
                  /map="16"
                  /clone="ct7-98d3"
                  /clone="bd3-6"
                  /clone_lib="CitbCJ7 mouse BAC library"
BASE COUNT   33314 a 31428 c 29695 g 32902 t
ORIGIN
Query Match      87.0%; Score 17.4; DB 12; Length 127339;
Best Local Similarity 94.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 gaacttcctaaagggagg 20
       ||||| ||||| |||||
Db      64435 GAAGTTCCTAAAGGGAGG 64453

RESULT      4
AC012297
LOCUS      AC012297 178964 bp DNA HTG 17-DEC-1999
DEFINITION Mus musculus clone RP23-78H9, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION  AC012297
VERSION     AC012297.2 GI:6598898
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 178964)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Mus musculus, clone RP23-78H9
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 178964)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferraira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Dec 17, 1999 this sequence version replaced gi:6094569.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RW/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L3207
            Center clone name: 78_H_9
            ----- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 152269 bases at least Q40
            Consensus quality: 168551 bases at least Q30
            Consensus quality: 174985 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 178964; sum-of-contigs
            Quality coverage: 5.7 in Q20 bases; agarose-fp
            Quality coverage: 5.8 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 1353: contig of 1353 bp in length
            * 1354 7574: contig of 6221 bp in length
            * 7575 12910: contig of 5336 bp in length
            * 12911 22836: contig of 9926 bp in length
            * 22837 29701: contig of 6865 bp in length
            * 29702 40566: contig of 10865 bp in length
            * 40567 51523: contig of 10957 bp in length
            * 51524 65303: contig of 13780 bp in length
            * 65304 83141: contig of 17838 bp in length
            * 83142 97592: contig of 14451 bp in length
            * 97593 122547: contig of 24955 bp in length
            * 122548 148570: contig of 26023 bp in length
            * 148571 178964: contig of 30394 bp in length.
            Location/Qualifiers
              1..178964
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP23-78H9"
                /clone_lib="RP23-78H9"
FEATURES     48859 a 43235 c 42064 g 44738 t 68 others
BASE COUNT
ORIGIN

```

```

Query Match      87.0%; Score 17.4; DB 43; Length 178964;
Best Local Similarity 94.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggag 19
|||||
Db 18694 GGAACCTCCCAAAGGAG 18712

RESULT 5
AC019079 AC019079 196435 bp DNA HTG 01-FEB-2000
LOCUS Homo sapiens clone RP11-334G8, WORKING DRAFT SEQUENCE, 13 unordered
DEFINITION pieces.
ACCESSION AC019079
VERSION AC019079.3 GI:6850740
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 196435)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 196435)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (30-DEC-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 1, 2000 this sequence version replaced gi:6693516.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0334G08
----- Summary Statistics -----
Sequencing vector: M13; 80%
Sequencing method: plasmid; 20%
Chemistry: Dye-primer ET; 80% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 185719 bases at least Q40
Consensus quality: 189576 bases at least Q30
Consensus quality: 192366 bases at least Q20
Insert size: 190000; agarose-fp
Quality coverage: 4.33 in Q20 bases; agarose-fp
Quality coverage: 4.19 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1748: contig of 1748 bp in length
* 1749 gap of unknown length
* 3695: contig of 1947 bp in length
* 3696 gap of unknown length
* 7126: contig of 3431 bp in length
* 7127 gap of unknown length
* 12216: contig of 5090 bp in length
* 12217 gap of unknown length
* 19115: contig of 6899 bp in length
* 19116 gap of unknown length
* 25360: contig of 6245 bp in length
* gap of unknown length

```

```

* 25361 38185: contig of 12825 bp in length
* gap of unknown length
* 38186 51067: contig of 12882 bp in length
* gap of unknown length
* 51068 67348: contig of 16281 bp in length
* gap of unknown length
* 67349 87975: contig of 20627 bp in length
* gap of unknown length
* 87976 108826: contig of 20851 bp in length
* gap of unknown length
* 108827 128076: contig of 19250 bp in length
* gap of unknown length
* 128077 196435: contig of 68359 bp in length.
FEATURES
    source
        1..196435
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-334G8"
BASE COUNT 58112 a 38351 c 37907 g 62054 t 11 others
ORIGIN

Query Match      87.0%; Score 17.4; DB 53; Length 196435;
Best Local Similarity 94.7%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggag 19
|||||
Db 180890 GGAACCTCCCTAAGGAG 180908

RESULT 6
AP000732/c 64714 bp DNA PLN 19-NOV-1999
LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone:F21A17,
DEFINITION complete sequence.
ACCESSION AP000732
VERSION AP000732.1 GI:6451837
KEYWORDS HTG
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:IGF BAC
clone:F21A17.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Nakamura,Y.
JOURNAL Structural Analysis of Arabidopsis thaliana Chromosome 3. III
REFERENCE 2 (bases 1 to 64714)
AUTHORS Nakamura,Y.
JOURNAL Direct Submission
TITLE Submitted (18-NOV-1999) to the DBJ/EMBL/GenBank databases.
UNIVERSITY Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/gene-s2/,
Tel:81-438-52-3935, Fax:81-438-52-3934)
FEATURES
    source
        1..64714
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /chromosome="3"
            /clone="F21A17"
            /clone_lib="IGF BAC"
BASE COUNT 20369 a 11946 c 12410 g 19989 t
ORIGIN

Query Match      85.0%; Score 17; DB 7; Length 64714;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 gaacttcctaaaggga 18          DNA          HTG          14-FEB-2000
      |||||
Db      12056 GAACCTCCCTAAAGGGA 12040
      |||||

RESULT      7
LOCUS      AC023483
DEFINITION Homo sapiens chromosome 3p clone RP11-453A3, WORKING DRAFT
SEQUENCE    AC023483
VERSION     AC023483.1 GI:6970723
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157644)
Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y.,
Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H.,
Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L.,
Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L.,
Feng, X., Yu, J. and Yang, H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 157644)
Wang, J., Hu, S., Dong, W., Zhang, X., Wang, X., Zhang, Y., Zhang, H.,
Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y.,
Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H.,
Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L.,
Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L.,
Feng, X., Yu, J. and Yang, H.
Direct Submission
Submitted (14-FEB-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 2296: contig of 2296 bp in length
* gap of unknown length
* 2297 4151: contig of 1855 bp in length
* gap of unknown length
* 4152 7388: contig of 3237 bp in length
* gap of unknown length
* 7389 12396: contig of 5008 bp in length
* gap of unknown length
* 12397 15967: contig of 3571 bp in length
* gap of unknown length
* 15968 20242: contig of 4275 bp in length
* gap of unknown length
* 20243 24704: contig of 4462 bp in length
* gap of unknown length
* 24705 30950: contig of 6246 bp in length
* gap of unknown length
* 30951 36291: contig of 5341 bp in length
* gap of unknown length
* 36292 40108: contig of 3817 bp in length
* gap of unknown length
* 40109 46878: contig of 6770 bp in length
* gap of unknown length
* 46879 55670: contig of 8792 bp in length
* gap of unknown length
* 55671 62531: contig of 6861 bp in length
* gap of unknown length
* 62532 70757: contig of 8226 bp in length

```

---

```

* 70758 84565: gap of unknown length
* contig of 13808 bp in length
* 84566 102424: gap of unknown length
* contig of 17859 bp in length
* 102425 118365: gap of unknown length
* contig of 15941 bp in length
* 118366 136073: gap of unknown length
* contig of 17708 bp in length
* 136074 157644: gap of unknown length
* contig of 21571 bp in length.
FEATURES
  source      Location/Qualifiers
 1..157644
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3p"
/clone="RP11-453A3"
BASE COUNT 48000 a 31264 c 30831 g 47545 t 4 others
ORIGIN

Query Match      85.0%; Score 17; DB 55; Length 157644;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggaacttcctaaaggg 17
      |||||
Db      33345 GGAACCTCCCTAAAGGG 33361

RESULT      8
LOCUS      AC016092/c
DEFINITION Homo sapiens clone RP11-25F1, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
ACCESSION    AC016092
VERSION      AC016092.2 GI:6573877
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165958)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 165958)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Beckerly, L., Boukhgaiter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 14, 1999 this sequence version replaced gi:6456222.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

Center project name: L4552  
 Center clone name: 25\_F\_1  
 ----- Summary Statistics -----  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 150475 bases at least Q40  
 Consensus quality: 160174 bases at least Q30  
 Consensus quality: 163440 bases at least Q20  
 Insert size: 170000; agarose-fp  
 Insert size: 165958; sum-of-contigs  
 Quality coverage: 5.2 in Q20 bases; agarose-fp  
 Quality coverage: 5.3 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 14 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2608: contig of 2608 bp in length  
 \* gap of unknown length  
 \* 2609 3894: contig of 1286 bp in length  
 \* gap of unknown length  
 \* 3895 6876: contig of 2982 bp in length  
 \* gap of unknown length  
 \* 6877 10815: contig of 3939 bp in length  
 \* gap of unknown length  
 \* 10816 15324: contig of 4509 bp in length  
 \* gap of unknown length  
 \* 15325 20970: contig of 5646 bp in length  
 \* gap of unknown length  
 \* 20971 27807: contig of 6837 bp in length  
 \* gap of unknown length  
 \* 27808 36301: contig of 8494 bp in length  
 \* gap of unknown length  
 \* 36302 51031: contig of 14730 bp in length  
 \* gap of unknown length  
 \* 51032 63828: contig of 12797 bp in length  
 \* gap of unknown length  
 \* 63829 80581: contig of 16753 bp in length  
 \* gap of unknown length  
 \* 80582 99556: contig of 18975 bp in length  
 \* gap of unknown length  
 \* 99557 121124: contig of 21568 bp in length  
 \* gap of unknown length  
 \* 121125 165958: contig of 44834 bp in length.  
 \* Location/Qualifiers  
 1. 165958  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-25F1"  
 /clone\_lib="RPC1-11 Human Male BAC"

# FEATURES

source  
 BASE COUNT 52147 a 32802 c 32547 g 48435 t 27 others  
 ORIGIN

Query Match 85.0%; Score 17; DB 43; Length 165958;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggg 17  
 |||||||||||||

Db 76935 GGAACCTCCCTAAAGGG 76919

RESULT 9  
 AC022800/c

LOCUS AC022800 176475 bp DNA 17-FEB-2000  
 DEFINITION Homo sapiens chromosome 1 clone RP11-336F23 map 1, WORKING DRAFT  
 SEQUENCE, 19 unordered pieces.

## ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AC022800  
 AC022800.2 GI:69844459  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 human  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

1 (bases 1 to 176475)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 1, clone RP11-336F23  
 Unpublished  
 2 (bases 1 to 176475)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Boda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 DeRellano,K., Dewar,K., Domino,M., Collins,S., Collymore,A., Cooke,P.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,  
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPheeters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Feb 17, 2000 this sequence version replaced gi:6922399.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 -----  
 Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 -----  
 Project Information  
 Center project name: L6049  
 Center clone name: 336\_F\_23  
 -----  
 Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 165291 bases at least Q40  
 Consensus quality: 171697 bases at least Q30  
 Consensus quality: 174526 bases at least Q20  
 Insert size: 176000; agarose-fp  
 Insert size: 176475; sum-of-contigs  
 Quality coverage: 4.0 in Q20 bases; agarose-fp  
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 19 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1339: contig of 1339 bp in length  
 \* gap of unknown length  
 \* 1340 4896: contig of 3557 bp in length  
 \* gap of unknown length  
 \* 4897 10931: contig of 6035 bp in length  
 \* gap of unknown length  
 \* 10932 17027: contig of 6096 bp in length  
 \* gap of unknown length  
 \* 17028 21477: contig of 4450 bp in length

```
*      21478      27370:      gap of unknown length
*      *      contig of 5893 bp in length
*      *      gap of unknown length
*      27371      33309:      contig of 5939 bp in length
*      *      gap of unknown length
*      33310      39351:      contig of 6042 bp in length
*      *      gap of unknown length
*      39352      49910:      contig of 10559 bp in length
*      *      gap of unknown length
*      49911      58253:      contig of 8343 bp in length
*      *      gap of unknown length
*      58254      66073:      contig of 7820 bp in length
*      *      gap of unknown length
*      66074      72593:      contig of 6520 bp in length
*      *      gap of unknown length
*      72594      81486:      contig of 8893 bp in length
*      *      gap of unknown length
*      81487      91215:      contig of 9729 bp in length
*      *      gap of unknown length
*      91216      103022:      contig of 11807 bp in length
*      *      gap of unknown length
*      103023      118782:      contig of 15760 bp in length
*      *      gap of unknown length
*      118783      137276:      contig of 18494 bp in length
*      *      gap of unknown length
*      137277      154651:      contig of 17375 bp in length
*      *      gap of unknown length
*      154652      176475:      contig of 21824 bp in length.
FEATURES
  source      Location/Qualifiers
1. 176475
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="1"
   /map="1"
   /clone="RP11-336F23"
   /clone_11b="RPC1-11 Human Male BAC"
BASE COUNT      52931 a 34068 c 34501 g 54964 t 11 others
ORIGIN

Query Match      85.0%; Score 17; DB 55; Length 176475;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 aacttcctaaaggag 19
|||||
Db 26894 AACTTCCTAAAGGAG 26878

RESULT 10
AC006560/c
LOCUS
DEFINITION      Homo sapiens BAC RPC111-792P18 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
ACCESSION      AC006560
VERSION
KEYWORDS
SOURCE
  ORGANISM      Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 202478)
  Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
  Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
  Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
  Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
  Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
  Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
  Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shih,C., Simon,M.,
  Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,
  Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
  Direct Submission
  JOURNAL
  TITLE
  JOURNAL
  2 (bases 1 to 202478)
  Worley,K.C.
  Direct Submission
  Submitted (12-FEB-1999) Molecular and Human Genetics, Baylor
  College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
  3 (bases 1 to 202478)
  Worley,K.C.
  Direct Submission
  Submitted (01-APR-1999) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  4 (bases 1 to 202478)
  Worley,K.C.
  Direct Submission
  Submitted (01-JUL-1999) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On Apr 2, 1999 this sequence version replaced gi:4521977.
  INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
  gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
  STSS are identified using ePCR (Genome Res. 7:541-550) searches
  of a local database that includes entries from dbSTS, GDB, and
  local mapping efforts.
  Repeats are identified using RepeatMasker (A. Smit and P. Green,
  unpublished.) for Human and Mouse sequences.
  Genes and Region of sequence similarity are identified by BLAST
  (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
  EST and cDNA sequences. Genes demonstrate at least two exons
  flanked by consensus splice sites that maintained sequence
  continuity across the splice junctions. Sequences that are not
  identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
html.

QUALSTAT-REPORT-----
----- Summary Statistics -----
Contig length:      202478
Phrap values in estimate:
Average error rate (BCM-Phrap estimate): 0.000107863
Fraction of Phrap values less than 4 : 0.0235058
Number of consensus changing edits: 1756
Number of N's in consensus : 0
----- Consensus changing edits -----
Position      Original+Context      Edited+Context
4435      agccaactcc(n)naaaccaca
4450      nccccaaac(c)aatgaagaac
4460      aatgaaga(a)c)ctccatcctt
4470      actccatcct(t)aatattctgt

REFERENCE
AUTHORS
  2 (bases 1 to 202478)
  Worley,K.C.
  Direct Submission
  Submitted (12-FEB-1999) Molecular and Human Genetics, Baylor
  College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
  3 (bases 1 to 202478)
  Worley,K.C.
  Direct Submission
  Submitted (01-APR-1999) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  4 (bases 1 to 202478)
  Worley,K.C.
  Direct Submission
  Submitted (01-JUL-1999) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On Apr 2, 1999 this sequence version replaced gi:4521977.
  INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
  gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
  STSS are identified using ePCR (Genome Res. 7:541-550) searches
  of a local database that includes entries from dbSTS, GDB, and
  local mapping efforts.
  Repeats are identified using RepeatMasker (A. Smit and P. Green,
  unpublished.) for Human and Mouse sequences.
  Genes and Region of sequence similarity are identified by BLAST
  (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
  EST and cDNA sequences. Genes demonstrate at least two exons
  flanked by consensus splice sites that maintained sequence
  continuity across the splice junctions. Sequences that are not
  identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
html.

QUALSTAT-REPORT-----
----- Summary Statistics -----
Contig length:      202478
Phrap values in estimate:
Average error rate (BCM-Phrap estimate): 0.000107863
Fraction of Phrap values less than 4 : 0.0235058
Number of consensus changing edits: 1756
Number of N's in consensus : 0
----- Consensus changing edits -----
Position      Original+Context      Edited+Context
4435      agccaactcc(n)naaaccaca
4450      nccccaaac(c)aatgaagaac
4460      aatgaaga(a)c)ctccatcctt
4470      actccatcct(t)aatattctgt
```

4481	aattctctgt(t)cctttagaac	atattctgtt(c)ctctagaacc	62576	tnannnnnt(n)nttnggnnnn	taagatagtt(g)tttgggatga
4490	tctcttaga(a)ccactccngt	tctcttagaa(c)caactctggt	62598	ttnggnnnn(n)nnnnnnnaa	tgtggtagag(a)gctctctaaa
4501	ccactccngt(a)agaaaactg	ccactctggt(g)aaactctgt	62603	nnnaaaaaa(n)tnnttttaa	tctaaaaaaa(t)ttcttttaa
4510	ttagaatac(t)atttagtgt	tagaaaactc(g)atttagtgt	62639	ataagatctc(n)ttcactgttc	ataagatctc(a)ttcactgttc
4521	gtattaggtt(t)ctctagaagg	tattagaggt(c)ctctagaagg	64371	agttaggttt(n)gttngngcaa	agttaggttt(t)gttngngcaa
4531	tctctaggtt(g)acagaactaa	ctctagaggg(a)cagaactaaa	64376	gtttngttgt(n)gcaannncn	gtttngttgt(t)gtcagaagcca
4541	acagaactaa(a)tagtatata	acagaactaa(t)tagtatata	64381	ngcaannnc(n)cntnnagnn	tgcagaagtc(a)ccttcagtcg
4551	atagatatata(t)atataatcc	tagtatata(t)atataatcc	64391	ncntnnagnn(n)nnnnnacta	accttcagtg(c)actgcactac
4561	tatatatac(c)attacatatc	attatatcc(t)attacatatc	98325	gagcactgca(n)ctagaagggg	gagcactgca(g)ctagaagggg
4571	tattacaata(t)ccatatatat	tattacaat(c)ccatatatat	98451	taagacagac(n)ctctgggcag	taagacagac(g)ctctgggcag
4581	tccatatata(t)atataatat	ccatatata(a)atataatat	98506	aaattatgca(n)antngtggg	aaattatgca(t)antngtggg
4592	atatataat(a)atataatat	tatatata(t)atataatat	98510	tatgcanaat(n)gttgggcctt	tatgcanaat(t)gttgggcctt
4601	tatatata(t)atataatat	atatataat(a)atataatat	108030	gagcacttaa(a)nnnnntttt	gagcacttaa(t)acatgcttta
4611	tatatata(t)atataataa	atatataat(a)atataataa	108042	annnttttaa(a)tnnnntttt	catgctttaa(t)atgatgtctt
4621	tatatata(a)tatgagttta	atatataaa(t)atgagtttat	108055	nnntttttt(n)ntttnnncc	gatgtctttt(a)ctttagtaacc
4631	ataagagttt(a)ttaagfatta	tatgaattta(t)taagfatta	108069	tnnnccntt(t)tnnttttna	tagtaccatt(c)ttcatgtgac
4642	taagtattta(a)ctcacatgat	taagtattaa(c)ctcacatgatc	108081	tnnttttna(n)ngnaacccc	ttcatgtgac(a)gtgaacccc
4651	aactcacagt(a)tcacaagctc	actcacatga(t)tcacaagctcc	108091	ngnaacccc(n)tnnnannnc	gtgaacccc(a)tcgatatgcca
4660	gatcacagct(t)cccatgagct	atcacagagt(c)ccatgagctg	108100	cntnnannn(n)cntttnttc	catgatagtc(c)actttatctc
4671	cccatgagct(g)ngnannaggg	ccatgagctg(a)ggagcaagg	108114	ttntccccc(n)nnnaccngn	ttatctcccc(a)cagaactcag
4681	gngnannagg(g)aancncnnn	aggagcaagg(a)aacccagctc	108122	ccnnnnnacc(n)nggntnncc	ccacagaacc(t)aggatagcc
4691	gaanccnnn(n)nnnnccnaa	aaacccagtc(t)gaatccaaa	108130	ccnnnggntn(n)ctttacacc	ccttagggtat(g)cttcaaccc
4703	nnnccnaaa(n)nnnanaant	agtcccaaaa(c)tgagaactt	108146	cacccctttt(c)tnnaannnn	acaccccttt(g)ttgaaatcag
4711	aannnnana(n)ntngnnncn	aactgagaag(c)ttggagtcgg	108152	tttttnnaa(n)nnnnnaaah	ttttgtgaa(t)catgaaaatg
4721	ntngnnnc(n)annnttnng	cttggagtc(g)atgtttgag	108161	annnnnaaa(n)nnnnnataa	atcatgaaa(t)gcagcataaa
4730	cnnnnttnn(n)ggncngnng	cgatgtttga(g)ggcaggaagg	108169	aannnnnaa(t)aaccgccaaa	aatgcagcat(a)aaccccaaaa
4742	gncngnagg(n)ncnngnng	gcaggaagg(t)ccagcgcaag	108180	aaacccaaa(a)aaantnnan	aaacccaaa(c)aaatctgata
4751	gncnnnnn(n)gnnaaannng	gtccagcgca(g)gagaagaatg	108190	aaaaantnna(n)gggnctttn	caaatctgat(a)gggtctcttg
4760	ngnnnaann(n)gnnggccagt	aggagaaga(g)gttgggccagt	108200	ngggnccttt(n)aatttggaa	agggtctctt(g)aatttggaa
10593	ctccctctc(n)nttctctccc	tctctctccc(c)cttctctccc	108210	naatttggaa(n)tgttttttn	gaatttggaa(t)gtgttttatg
10595	tctctctccc(n)cttctctccc	ctctctctcc(c)cttctctccc	108220	cttgcctccc(n)ntttctctn	tggttttat(a)ctttctctta
10657	ctctgcctcc(n)tcctctcttc	ctctgcctcc(c)tcctctcttc	108230	nnnttctct(n)ccnnnttccc	gctttctctt(a)cttgccttcc
10736	ctcctctccc(n)cccctctccc	ctcctctccc(c)tcctctctcc			
10972	ctcccctccc(n)tcctctctcc	ctcccctccc(c)tcctctctcc			
13942	ctcccctctcc(a)cccctctcgc	ctcccctccc(c)cccctctcgc			
26093	ctctatttta(c)gcgaagctt	ctctatttta(g)gcgaagcttc			
26111	cttctgttcc(n)ccctgtgagc	cttctgttcc(t)ccctgtgagc			
26165	agattctaaa(n)ccctgtgtg	agattctaaa(a)ccctgtgtg			
26188	ccattgagag(n)tcattgggca	ccattgagag(a)tcattgggca			
26237	ggaactagct(n)ncctnccct	ggaactagct(c)tcctnccct			
26238	gaactagctn(n)ccnccctct	gaactagctt(c)ccnccctct			
26242	tagctnnccct(n)cccctctcnn	tagctctcct(t)cccctctcgc			
26251	tncccctccc(n)nggcntttn	tncccctccc(g)caggctctca			
26260	cnngngctt(n)ngncennnt	cgagggcttc(c)aggcctcatt			
26273	ncennnttt(n)agctctgggg	gcctaatctt(g)agctctgggg			
26292	ggtttgagac(n)ccgttctcgg	ggtttgagac(a)ccgttctcgg			
26313	atgttgaag(n)ccnnnnnggg	atgttgaag(t)ccacagaggg			
26320	aagncnnnn(n)ggggnggcaa	aagtcacacag(a)ggggngcaa			
26325	cnnnngggg(n)ggcaagacac	cacagagggg(a)ggcaagacac			
51215	tcgtctctt(n)tcgtgttga	tcgtctctt(t)tcgtgttga			
56757	ttttatgct(n)ataatcactt	ttttatgct(t)ataatcactt			
62235	ttgcagggaa(n)taaaagcgag	ttgcagggaa(g)taaaagcgag			
62241	gcgagcagtt(n)aacgttttaa	gcgagcagtt(g)aacgttttaa			
62264	atttctggg(g)aagttggagg	atttctggg(a)aagttggagg			
62301	ggaaactgtt(n)ngggnnngg	ggaaactgtt(g)aggagatggg			
62312	agggnnggg(n)ncnntttn	agggngctgg(a)ccagatcttg			
62322	ncnntttt(n)ngnngnngn	acagagttt(g)aggagaggtt			
62332	ngnngnnng(n)tnanntnna	gaggagggg(t)taagttcag			
62342	tnannttna(n)nnngngnnc	ttgaagttta(g)ctagggtccc			
62353	nnngngnnc(n)nnnnctnn	ctagggttcc(a)gataaccaga			
62363	nnnnnnnn(n)annctntnn	agatatccag(a)agactcttgc			
62373	nannntnn(n)gngngcnng	aagacctct(g)gagggccagg			
62385	ngngcnngg(n)nnngnnaht	agggccagg(a)cagggaattg			
62395	nnnggnant(n)gmnnaang	acagggaatt(g)gagataaagg			
62404	tnngnnnaa(n)ggaggttaga	tggaagataa(g)ggaggttaga			
62421	tagaagtagg(n)ggagaaggag	tagaagtagg(t)ggagaaggag			
62463	ttggcttttg(n)ttcaattct	ttggctttt(g)ttcaattct			
62518	gctagcagtt(n)caacttntg	gctagcagtt(a)caacttctg			
62526	ttncactctt(n)tggtccnnn	ttncactctt(c)tggtccctct			
62534	ttngtggccc(n)nttttnant	ttctgtgccc(t)ctttctacca			
62545	nttttnant(n)nnnnntnnc	cttttcaatt(a)gactctacca			
62555	nnnnntnnc(n)nnngnncnn	agactctacc(a)tagtatacct			
62565	nnngnncnn(n)tnannnnnt	atagggtacce(t)taagatggtt			
<hr/>					
Query Match 85.0%; Score 17; DB 40; Length 202478;					
Best Local Similarity 100.0%; Pred. No. 38;					
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	4	actctccataaaggag 20			
Db	135659	ACTTCCTAAAGGAGG 135643			
<hr/>					
RESULT 11					
LOCUS	AB000879	39436 bp DNA PRI 20-NOV-1999			
DEFINITION	Homo sapiens genomic DNA for centromeric end of MHC class I region on chromosome 6, cosmid clone: TY3A9, complete sequence.				
ACCESSION	AB000879				
VERSION	AB000879.1	GI:3021695			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (haplotype:A2 B62 CW10 DR4) adult male blood clone:745D12 sub_clone:cosmid clone:TY3A9.				
ORGANISM	Homo sapiens				
REFERENCE	Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Shina,T., Tamiya,G., Oka,A., Yamagata,T., Yamagata,N., Kikkawa,E., Goto,K., Mizuki,N., Watanabe,K., Fukuzumi,Y., Taguchi,S., Sugawara,C., Ono,A., Chen,L., Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Kimura,M. and Inoko,H.				
TITLE	Nucleotide sequencing analysis of the 146-kilobase segment around the IkbL and MICA genes at the centromeric end of the HLA class I region				
JOURNAL	Genomics 47 (3), 372-382 (1998)				
MEDLINE	98149985				
REFERENCE	2 (bases 1 to 39436)				
AUTHORS	Shina.T.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi Shina, Tokai University School of Medicine, Department of Molecular Life Sciences; Bohseidai, Isehara, Kanagawa 259-11, Japan (E-mail:tshina@is.icc.u-tokai.ac.jp, Tel.0463-93-1121,				

Fax:0463-94-8884)

FEATURES  
source  
1. .39436  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="BOLETH"  
/cell\_type="immunoresponce cell"  
/chromosome="6"  
/clone="745D12"  
/clone\_lib="CEPH YAC"  
/dev\_stage="adult"  
/haplotype="A2 B62 CW10 DR4"  
/map="6p21.3"  
/sex="male"  
/sub\_clone="cosmid clone:TY3A9"  
/tissue\_type="blood"  
BASE COUNT 10807 a 9318 c 8837 g 10474 t  
ORIGIN

Query Match 84.0% Score 16.8; DB 9; Length 39436;  
Best Local Similarity 90.0%; Pred. No. 52;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggaacttcctaaaggagg 20  
|||||  
Db 1393 GGAACCTCCCTTAGGGGAG 1374

RESULT 12  
AC004184/c 44710 bp DNA PRI 08-DEC-1998  
LOCUS  
DEFINITION Homo sapiens clone UWGC:R3A from 6p21, complete sequence.  
AC004184  
VERSION AC004184.1 GI:3980476  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Guillaudeau,T., Janer,M.M., Wong,G.K.-S., Spies,T. and  
Geraghty,D.E.  
The complete genomic sequence of 424,015 bp at the centromeric end  
of the HLA class I region: gene content and polymorphism  
Proc. Nat. Acad. Sci., USA (1998) in press  
Fred Hutchinson Cancer Research Center  
The Clinical Research Division  
1100 Fairview Ave. N., P.O. Box 19024  
Seattle, WA 98109-1024  
2 (bases 1 to 44710)  
Geraghty,D.E. and Olson,M.V.  
Direct Submission  
Submitted (23-FEB-1998) Human Genome Center, University of  
Washington, Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 44710)  
Geraghty,D.E. and Olson,M.V.  
Direct Submission  
Submitted (08-DEC-1998) Human Genome Center, University of  
Washington, Box 352145, Seattle, WA 98195, USA  
University of Washington Human Genome Center  
Box 352145, Seattle, WA 98195  
Contact: Daniel E. Geraghty (efhcr.org)  
On Dec 8, 1998 this sequence version replaced gi:2905876.  
Overlapping Sequences:  
5': UWGC:R15A (Genbank Accession: AC004183)  
3': UWGC:R13C (Genbank Accession: AC004215)  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than

1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

Double stranded (DS) coverage: 90.1%  
DS or two chemistry coverage: 99.7%  
Single stranded regions: 1  
-----  
Sequence Validation:  
This sequence has been validated by multiple enzyme digestion  
of the clone. The fingerprint comparison with sequence-predicted  
fragments is given below.  
Small fragments below a variable cutoff (approximately 400-600bp)  
are not mapped and hence do not appear in the table. There are no  
significant remaining discrepancies between the experimental and  
predicted values. Uniquely ordered fragment groups are separated  
by dashed lines.

	BglII	Seq	HindIII	Map	Seq	Map	NsiI
-----	-----	-----	-----	-----	-----	-----	-----
5118.00	5043.00	8808.00	8617.00	1727.00	1729.00		
3909.00	3843.00	845.00	816.00	7139.00	7025.00		
5118.00	5074.00	6557.00	6394.00	3686.00	3613.00		
1868.00	1860.00	7550.00	7331.00	2080.00	2072.00		
8369.00	8234.00	15465.00	15651.00	930.00	916.00		
804.00	790.00			675.00	669.00		
1341.00	1336.00			5695.00	5633.00		
4474.00	4414.00			7139.00	7168.00		
8521.00	8471.00			4151.00	4165.00		
2797.00	2911.00						
-----	-----	-----	-----	-----	-----	-----	-----
FEATURES	Location/Qualifiers						
source	1. .44710						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/chromosome="6"						
	/map="6p21"						
	/clone="UWGC:R3A"						
	/cell_line="b-lymphoblastoid cell line Mann"						
	/clone_lib="T. Spies HLA Cosmid Library"						
	/note="Spies, T., Breshahan, M. & Strominger, J.L.: Human						
	major histocompatibility complex contains a minimum of 19						
	genes between the complement cluster and HLA-B. Proc Natl						
	Acad Sci U S A 86:8955-8, 1989"						
repeat_region	445. .597						
	/rpt_family="LTR8"						
repeat_region	complement(818. .1103)						
	/rpt_family="Alu"						
repeat_region	1118. .1458						
	/rpt_family="LTR8"						
repeat_region	complement(2242. .2476)						
	/rpt_family="Alu"						
repeat_region	complement(5735. .5852)						
	/rpt_family="MERS"						
repeat_region	8075. .8372						
	/rpt_family="Alu"						
repeat_region	complement(9483. .9775)						
	/rpt_family="Alu"						
repeat_region	10587. .10883						
	/rpt_family="Alu"						
repeat_region	15345. .15450						
	/rpt_family="Alu"						
repeat_region	16661. .16972						



\* 19747 20604: contig of 858 bp in length  
\* gap of unknown length  
\* 20605 21476: contig of 872 bp in length  
\* gap of unknown length  
\* 21477 22353: contig of 877 bp in length  
\* gap of unknown length  
\* 22354 23219: contig of 866 bp in length  
\* gap of unknown length  
\* 23220 24111: contig of 892 bp in length  
\* gap of unknown length  
\* 24112 25101: contig of 990 bp in length  
\* gap of unknown length  
\* 25102 25976: contig of 875 bp in length  
\* gap of unknown length  
\* 25977 26856: contig of 880 bp in length  
\* gap of unknown length  
\* 26857 27724: contig of 868 bp in length  
\* gap of unknown length  
\* 27725 28604: contig of 880 bp in length  
\* gap of unknown length  
\* 28605 29481: contig of 877 bp in length  
\* gap of unknown length  
\* 29482 30358: contig of 877 bp in length  
\* gap of unknown length  
\* 30359 31302: contig of 944 bp in length  
\* gap of unknown length  
\* 31303 32162: contig of 860 bp in length  
\* gap of unknown length  
\* 32163 33036: contig of 874 bp in length  
\* gap of unknown length  
\* 33037 33873: contig of 837 bp in length  
\* gap of unknown length  
\* 33874 34752: contig of 879 bp in length  
\* gap of unknown length  
\* 34753 35638: contig of 886 bp in length  
\* gap of unknown length  
\* 35639 36513: contig of 875 bp in length  
\* gap of unknown length  
\* 36514 37369: contig of 856 bp in length  
\* gap of unknown length  
\* 37370 38271: contig of 902 bp in length  
\* gap of unknown length  
\* 38272 39161: contig of 890 bp in length  
\* gap of unknown length  
\* 39162 40040: contig of 879 bp in length  
\* gap of unknown length  
\* 40041 40897: contig of 857 bp in length  
\* gap of unknown length  
\* 40898 41884: contig of 987 bp in length  
\* gap of unknown length  
\* 41885 42745: contig of 861 bp in length  
\* gap of unknown length  
\* 42746 43627: contig of 882 bp in length  
\* gap of unknown length  
\* 43628 44497: contig of 870 bp in length  
\* gap of unknown length  
\* 44498 45365: contig of 868 bp in length  
\* gap of unknown length  
\* 45366 46219: contig of 854 bp in length  
\* gap of unknown length  
\* 46220 47090: contig of 871 bp in length  
\* gap of unknown length  
\* 47091 47970: contig of 880 bp in length  
\* gap of unknown length  
\* 47971 48838: contig of 868 bp in length  
\* gap of unknown length  
\* 48839 49738: contig of 900 bp in length  
\* gap of unknown length  
\* 49739 50656: contig of 918 bp in length  
\* gap of unknown length  
\* 50657 51554: contig of 898 bp in length  
\* gap of unknown length  
\* 51555 52420: contig of 866 bp in length

\* 52421 53288: contig of 868 bp in length  
\* gap of unknown length  
\* 53289 54222: contig of 934 bp in length  
\* gap of unknown length  
\* 54223 55117: contig of 895 bp in length  
\* gap of unknown length  
\* 55118 56022: contig of 905 bp in length  
\* gap of unknown length  
\* 56023 56889: contig of 867 bp in length  
\* gap of unknown length  
\* 56890 57854: contig of 965 bp in length  
\* gap of unknown length  
\* 57855 58757: contig of 903 bp in length  
\* gap of unknown length  
\* 58758 59659: contig of 902 bp in length  
\* gap of unknown length  
\* 59660 60582: contig of 923 bp in length  
\* gap of unknown length  
\* 60583 61461: contig of 879 bp in length  
\* gap of unknown length  
\* 61462 62454: contig of 993 bp in length  
\* gap of unknown length  
\* 62455 63429: contig of 975 bp in length  
\* gap of unknown length  
\* 63430 64406: contig of 977 bp in length.  
\* Location/Qualifiers  
\* 1. 64406  
\* /organism="Homo sapiens"  
\* /db\_xref="taxon:9606"  
\* /chromosome="18"  
\* /map="18"  
\* /clone="57\_H\_1"  
\* /clone\_lib="RPC1-11 Human Male BAC"  
\* BASE COUNT 17340 a 14550 c 14451 g 17597 t 468 others  
\* ORIGIN

# FEATURES

Query Match 84.0%; Score 16.8; DB 41; Length 64406;  
Best Local Similarity 90.0%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggagg 20  
|||||

Db 13417 GGGATTCCCTAAAGGAGG 13436

# RESULT 14

AP000506/c

LOCUS AP000506 100000 bp DNA PRI 28-SEP-1999  
DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,  
section 5/20.

ACCESSION AP000506

VERSION AP000506.1 GI:5926693

KEYWORDS

SOURCE Homo sapiens DNA.

# ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

# REFERENCE

1 (sites)

Shiina,S., Tamiya,G., Oka,A. and Inoko,H.

Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region

PUBLISHED ONLY in database (1999) In press

2 (bases 1 to 100000)

Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.

# TITLE

Direct Submission

# JOURNAL

Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika

Hirakawa, Japan Science and Technology Corporation (JST), Advanced

Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,

Japan (E-mail:mika@tokyo.jst.go.jp,

URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,

Fax:81-3-5214-8470)

# COMMENT

This sequence is conducted by Tokai University as a JST sequencing

Team.  
Principal Investigator: Hidetoshi Inoko Ph.D  
Phone: +81-463-93-1121, Fax: +81-463-94-8884,  
The sequence is submitted by Human Genome Sequencing in ALIS  
project of JST  
Japan Science and Technology Corporation (JST)  
5-3, Yonbancyo, Chiyoda-ku, Tokyo, 102-0081 Japan  
For further information about this sequences, please visit our  
sequence archive web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>) or send email to [webmaster@www-alis.tokyo.jst.go.jp](mailto:webmaster@www-alis.tokyo.jst.go.jp).

## FEATURES

source  
1..100000  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21.3"  
STS  
complement(13599..13701)  
/note="stSG1314; The location is between each flanking site  
of PCR primers."  
/db\_xref="GDB:4562247"  
gene  
complement(32881..45810)  
/gene="MIC9"  
misc\_feature  
complement(32881..45810)  
/gene="MICB"  
/note="match MICB (U65416) with account of Intervene  
sequences."  
misc\_feature  
complement(78169..80801)  
/gene="HS.1845"  
/note="match HS.1845 (L06175) with account of intervene  
sequences."  
gene  
complement(78169..80801)  
/gene="HS.1845"  
STS  
complement(78405..78633)  
/gene="HS.1845"  
/note="RH17599; The location is between each flanking site  
of PCR primers."  
/db\_xref="GDB:4571965"  
BASE COUNT 26491 a 23361 c 23483 g 26665 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 100000;  
Best Local Similarity 90.0%; Pred. No. 50;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ggaacttcctaaaggagg 20  
||||||| |||||  
Db 67967 GGAACCTCCCTTAAGGGAAG 67948

RESULT 15  
AC018412/c  
LOCUS AC018412 135953 bp DNA HTG 29-DEC-1999  
DEFINITION Homo sapiens clone RP11-1718, WORKING DRAFT SEQUENCE, 13 unordered  
pieces.  
ACCESSION AC018412  
VERSION AC018412.2 GI:6646811  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 135953)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens, clone RP11-1718  
JOURNAL Unpublished  
REFERENCE  
2 (bases 1 to 135953)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boguslavsky,L., Boukigalter,B., Brown,A., Castle,A., Colangelo,M.,  
Collins,S., Collumore,A., Cooke,P., DeArellano,K., Dewar,K.,  
Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W.,  
Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,

Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,  
Karatas,A., Klein,J., Landers,T., Lehoczy,J., Lieu,C., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,  
Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

## COMMENT

Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 29, 1999 this sequence version replaced gi:6554552.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L3648

Center clone name: 17\_I-8

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 126643 bases at least Q40

Consensus quality: 131641 bases at least Q30

Consensus quality: 133760 bases at least Q20

Insert size: 139000; agarose-fp

Insert size: 135953; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; agarose-fp

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 470: contig of 470 bp in length  
\* gap of unknown length  
\* 471 2070: contig of 1600 bp in length  
\* gap of unknown length  
\* 2071 4004: contig of 1934 bp in length  
\* gap of unknown length  
\* 4005 7726: contig of 3722 bp in length  
\* gap of unknown length  
\* 7727 13560: contig of 5834 bp in length  
\* gap of unknown length  
\* 13561 18873: contig of 5313 bp in length  
\* gap of unknown length  
\* 18874 25091: contig of 6218 bp in length  
\* gap of unknown length  
\* 25092 32734: contig of 7643 bp in length  
\* gap of unknown length  
\* 32735 39860: contig of 7126 bp in length  
\* gap of unknown length  
\* 39861 48226: contig of 8366 bp in length  
\* gap of unknown length  
\* 48227 58275: contig of 10049 bp in length  
\* gap of unknown length  
\* 58276 82056: contig of 23781 bp in length  
\* gap of unknown length  
\* 82057 135953: contig of 53897 bp in length.

## FEATURES

## source

Location/Qualifiers  
1..135953  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="RP11-1718"  
/clone\_lib="RPC1-11 Human Male BAC"  
BASE COUNT 40561 a 28248 c 28138 g 39005 t 1 others  
ORIGIN

Query Match 84.0%; Score 16.8; DB 44; Length 135953;  
Best Local Similarity 90.0%; Pred. No. 50;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ggaacttcctaaaggagg 20  
||||| ||||| ||||| |||||  
Db 93676 GGAACATCCCCCAAGGAGG 93657

Search completed: May 23, 2000, 09:40:53  
Job time: 8829 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:42:07 ; Search time 88.1 Seconds  
(without alignments)  
56.797 Million cell updates/sec

Title: US-08-945-805-3  
Perfect score: 20  
Sequence: 1 ggaactccctaaaggagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.4	77.0	588	T68207	Helicobacter pylori
C 2	15.4	77.0	651	X14510	H. pylori GHPO 124
C 3	15.4	77.0	3560	Q40312	Helicobacter pylori
C 4	14.8	74.0	1144	V18014	Pinus radiata cone
C 5	14.8	74.0	110000	V21209_13	Continuation (14 o
C 6	14.4	72.0	468	X30913	Streptococcus pneu
C 7	14.4	72.0	744	V84493	Human secreted pro
C 8	14.4	72.0	780	N80975	Probe pCS.7 for sc
C 9	14.4	72.0	1041	T68000	H. pylori inner me
C 10	14.4	72.0	1146	V25046	H. pylori cytoplas
C 11	14.4	72.0	1158	V24720	H. pylori ORF 06ep
C 12	14.4	72.0	1350	X14148	H. pylori GHPO 422
C 13	14.4	72.0	1404	Q43719	Sequence which hyb
C 14	14.4	72.0	1404	X16664	Bovine Y chromosom
C 15	14.4	72.0	3134	Q98950	b98 insert encodin
C 16	14.4	72.0	4488	Q51426	Human FACC cDNA cl
C 17	14.4	72.0	4567	V33945	Fanconi anaemia cl
C 18	14.4	72.0	19932	T46159	CagI locus. Helico
C 19	14.2	71.0	44	R80370	Heteropolymer olig
C 20	14.2	71.0	372	Q60269	Human brain expres
C 21	14.2	71.0	1065	T67710	H. pylori cell env
C 22	14.2	71.0	1074	T67988	H. pylori cell env
C 23	14.2	71.0	1512	V01597	Tobacco lactase cl
C 24	14.2	71.0	1608	T40731	Endonuclease 2 enc
C 25	14.2	71.0	2169	X14202	H. pylori GHPO 992
C 26	14.2	71.0	5340	T10258	DNA polymerase gen
C 27	14.2	71.0	5340	T1296	KOD1 thermostable
C 28	14.2	71.0	5342	T28360	DNA polymerase gen
C 29	14.2	71.0	5342	T28360	DNA polymerase gen
C 30	14.2	71.0	5342	T14694	DNA polymerase. En
C 31	14.2	71.0	5342	T14694	DNA polymerase. En
C 32	14.2	71.0	5342	V14525	Full length heat-r
C 33	14.2	71.0	5342	V14525	Full length heat-r
C 34	14.2	71.0	22481	T11658	PEDF full length s

C 35	13.8	69.0	273	1	T19188	Human gene signatu
C 36	13.8	69.0	1003	1	T40041	presenilin-1 exon
C 37	13.8	69.0	1003	1	T99670	Human presenilin-1
C 38	13.8	69.0	1152	1	X14434	H. pylori GHPO 800
C 39	13.8	69.0	1751	1	N71021	Sequence encoding
C 40	13.8	69.0	4080	1	T30652	P. gingivalis haem
C 41	13.8	69.0	4080	1	V58873	Haemagglutinin pro
C 42	13.8	69.0	4998	1	N81114	Non-A, non-B hepat
C 43	13.8	69.0	5019	1	X20583	Polynucleotide seq
C 44	13.8	69.0	6000	1	T78851	Lys-specific thiol
C 45	13.8	69.0	6241	1	T30655	P. gingivalis haem

ALIGNMENTS

RESULT 1

T68207/c

ID T68207 standard; DNA; 588 BP.

AC T68207;

DT 21-JUL-1997 (first entry)

DE Helicobacter pylori-derived open reading frame.

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;

KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;

KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.

OS Helicobacter pylori.

FH Key Location/Qualifiers

FT cds 1..588

FT /\*tag= a

FT /note= "no stop codon given"

PN WO9640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR ) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI: 97-052306/05.

DR P-PSDB: W20954.

PT Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

PS Disclosure; Page 963; 1481pp; English.

CC The present sequence encodes a Helicobacter pylori-derived protein

CC of unknown function (no further details given in the specification).

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To

CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.

SQ Sequence 588 BP; 130 A; 123 C; 139 G; 196 T;

Query Match 77.0%; Score 15.4; DB 1; Length 588;

Best Local Similarity 94.1%; Pred. No. 26;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaactccctaaaggga 18

|||||

Db 510 GAATTTCCCTAAGGGA 494

RESULT 2

X14510/c

ID X14510 standard; DNA; 651 BP.

```

AC X14510;
DT 31-MAR-1999 (first entry)
DE H. pylori GHPO 1246 gene.
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 43..630
FT /*tag= a
FT WO9843478-A1.
PN 08-OCT-1998.
PD 01-APR-1998; U06371.
PF 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INRM ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI; 98-542293/46.
DR P-PSDB; W98791.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 1721-1722; 2054pp; English.
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 651 BP; 148 A; 123 C; 156 G; 224 T;

Query Match 77.0%; Score 15.4; DB 1; Length 651;
Best Local Similarity 94.1%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaacttcctccctaaaggga 18
Db 549 GAATTTCCCTAAAGGGA 533

RESULT 3
Q40312/c
ID Q40312 standard; DNA; 3560 BP.
AC Q40312;
DT 03-AUG-1993 (first entry)
DE Helicobacter pylori urease operon.
KW ureE; ureF; ureG; ureH; ureI; stomach; gastric mucosa; gastric ulcer;
KW duodenal ulcer; acid tolerance; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT rbs 198..201
FT /*tag= a
FT /*standard_name= Shine_Dalgarno
FT cds 212..799
FT /*tag= b
FT /*label= ureI
FT cds 801..1313
FT /*tag= c
FT /*label= ureE
FT rbs 1313..1316
FT /*tag= d
FT /*standard_name= Shine_Dalgarno
FT cds 1325..2095
FT /*tag= e
FT /*label= ureF
FT misc_feature 1619..1634
FT /*tag= f
FT /*note= "sigma-54 promoter-like sequence"
FT cds 2123..2722

```

```

FT /*tag= g
FT /*label= ureG
FT rbs 2713..2717
FT /*tag= h
FT /*standard_name= Shine_Dalgarno
FT cds 2722..3519
FT /*tag= i
FT /*label= ureH
FT terminator 56..89
FT /*tag= j
FT /*note= "rho-independent hairpin"
FT rbs 2111..2116
FT /*tag= k
FT /*standard_name= Shine_Dalgarno
FT terminator 3528..3560
FT /*tag= l
FT /*note= "rho-independent hairpin"
FT WO9307273-A.
PN 15-APR-1993.
PD 02-OCT-1992; F00921.
PF 03-OCT-1991; FR-012198.
PR (INRM ) INSERM INST NAT SANTE & RECH MED.
PA (INRM ) INST PASTEUR.
PI Cussac V, Ferero R, Labigne A;
DR WPI; 93-134462/16.
DR P-PSDB; R34395, R34396, R34397, R34398, R34399.
PT Helicobacter pylori genes useful in diagnosis, vaccines and
PT treatment - necessary for the regulation and maturation of urease
PS Claim 7-11; Fig 4; 94pp; French.
CC The five new urease genes ure E, F, G, H and I were identified by
CC deletion studies in E.coli where all 5 were found to be necessary
CC for functional urease expression. They encode proteins of
CC estimated mol. wt. 19.5, 28.6, 21.7, 29.6 and 21.7kd, respectively.
CC Recombinant H.pylori having a mutation in at least one of the 5 new
CC genes or in ureA or ureB which renders them urease-negative or
CC urease-depleted (attenuated strains) can be used as immunogens to
CC protect against H.pylori infection.
SQ Sequence 3560 BP; 1059 A; 699 C; 776 G; 1026 T;

Query Match 77.0%; Score 15.4; DB 1; Length 3560;
Best Local Similarity 94.1%; Pred. No. 29;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaacttcctccctaaaggga 18
Db 718 GAATTTCCCTAAAGGGA 702

RESULT 4
V18014
ID V18014 standard; DNA; 1144 BP.
AC V18014;
DT 28-AUG-1998 (first entry)
DE Pinus radiata cone-specific pRMADS3 cDNA.
KW pRMADS3 gene; cone-specific gene; radiata pine; conifer;
KW forest tree; timber; pulp; fibre; transgenic plant; ss.
OS Pinus radiata.
FH Key Location/Qualifiers
FT CDS 1..786
FT /*tag= a
FT WO9813503-A1.
PN 02-APR-1998.
PD 23-SEP-1997; AU0625.
PR 24-FEB-1997; US-804879.
PR 23-SEP-1996; US-717971.
PR 04-OCT-1996; AU-002756.
PR 13-FEB-1997; AU-005092.
PA (EBIN-) FB INVESTMENTS PTY LTD.
PI Teasdale RD;
DR WPI; 98-230712/20.
DR P-PSDB; W48623.
PT Method of enhancing vegetative growth in plant - comprises use of

```

PT expression cassette containing heterologous coding region expressing  
 PT product which aborts development

PS Disclosure; Fig 13; 96pp; English.  
 CC This cDNA clone corresponds to a Pinus radiata cone-specific gene,  
 CC designated PrMADS3, whose expression is substantially restricted to  
 CC reproductive organ primordium tissues. 5 Cone-specific genes,  
 CC PrMADS1, 2 and 3, PrFL1 and PrCon1 (see also V18011-18), displaying  
 CC strong homology to Arabidopsis thaliana and Antirrhinum majus  
 CC floral meristem and organ identity genes were identified in a Pinus  
 CC radiata cDNA library prepared from immature female and male cones.  
 CC PrMADS1, PrMADS2 and PrMADS3 are MADS-box genes. The promoter  
 CC regions of these cone-specific genes, as well as promoter regions  
 CC (see V18019-22) of novel MADS-box genes of Eucalyptus can be used  
 CC in a novel method of enhancing vegetative growth in a plant. This  
 CC involves: identifying a gene having a tissue-specific promoter  
 CC expressed during the development of both male and female plant  
 CC reproductive structures; constructing an expression cassette  
 CC comprising a heterologous coding region capable of expressing a  
 CC product (e.g. barnase) which aborts the development under the  
 CC expression control of the promoter; (c) transforming plants, and  
 CC (d) selecting and vegetatively propagating the transformants.  
 CC Preferably, the plant is a tree for timber, pulp or fibre production  
 CC in which poor or absent expression of reproductive structures may  
 CC result in increased vegetative growth of the valuable material.  
 SQ Sequence 1144 BP; 377 A; 232 C; 253 G; 282 T;

Query Match 74.0%; Score 14.8; DB 1; Length 1144;  
 Best Local Similarity 88.9%; Pred. No. 56;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggaacttcctcctaaaggga 18  
 ||||| ||| |||||  
 DB 453 GGAACCTCCCAAGGGA 470

## RESULT 5

V21209\_13/c  
 Continuation (14 of 17) of V21209 from base 1300001 (Methanococcus jannaschii circular  
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 74.0%; Score 14.8; DB 1; Length 110000;  
 Best Local Similarity 88.9%; Pred. No. 76;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggaacttcctcctaaaggga 18  
 ||||| ||||| |||  
 DB 81732 GGAACCTCCCTACAGAGA 81715

## RESULT 6

X30913  
 ID X30913 standard; DNA; 468 BP.  
 AC X30913;

DT 20-MAY-1999 (first entry)  
 DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:190.  
 KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
 OS Streptococcal infection; pneumococcal; ss.  
 PN Streptococcus pneumoniae.  
 WO9737026-A1.

PD 09-OCT-1997.  
 PF 01-APR-1997; U05306.  
 PR 22-AUG-1996; US-025788.  
 PR 02-APR-1996; US-014690.  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Stodola RK;  
 DR WPI: 97-503111/46.

DR P-PSDB; Y11331.  
 PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
 PT vaccines, drug screening, etc

PS Claim 5; Page 166; 354pp; English.  
 CC X30724 to X30946 represent genomic DNA sequences isolated from  
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
 CC encode the novel proteins given in Y1114 to Y11367. The proteins,  
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
 CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.  
 SQ Sequence 468 BP; 156 A; 58 C; 87 G; 167 T;

Query Match 72.0%; Score 14.4; DB 1; Length 468;  
 Best Local Similarity 93.8%; Pred. No. 84;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaacttcctcctaaagg 17  
 ||||| ||||| |||  
 DB 331 GAACTTCCTCTAAAGAG 346

## RESULT 7

V84493  
 ID V84493 standard; DNA; 744 BP.

AC V84493;  
 DT 01-MAR-1999 (first entry)  
 DE Human secreted protein gene 83 clone HNGIN60.  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 PN W09854963-A2.  
 PD 10-DEC-1998.  
 PF 04-JUN-1998; U11422.  
 PR 18-DEC-1997; US-070923.  
 PR 06-JUN-1997; US-048877.  
 PR 06-JUN-1997; US-048881.  
 PR 06-JUN-1997; US-048884.  
 PR 06-JUN-1997; US-048893.  
 PR 06-JUN-1997; US-048896.  
 PR 06-JUN-1997; US-048899.  
 PR 06-JUN-1997; US-048915.  
 PR 06-JUN-1997; US-048949.  
 PR 06-JUN-1997; US-048964.  
 PR 06-JUN-1997; US-048972.  
 PR 06-JUN-1997; US-049020.  
 PR 06-JUN-1997; US-049375.  
 PR 05-SEP-1997; US-057628.  
 PR 05-SEP-1997; US-057635.

```

PR 05-SEP-1997; US-057644.
PR 05-SEP-1997; US-057647.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057661.
PR 05-SEP-1997; US-057667.
PR 05-SEP-1997; US-057761.
PR 05-SEP-1997; US-057764.
PR 05-SEP-1997; US-057770.
PR 05-SEP-1997; US-057775.
PR 05-SEP-1997; US-057778.
PR 06-JUN-1997; US-048875.
PR 06-JUN-1997; US-048878.
PR 06-JUN-1997; US-048882.
PR 06-JUN-1997; US-048885.
PR 06-JUN-1997; US-048894.
PR 06-JUN-1997; US-048897.
PR 06-JUN-1997; US-048900.
PR 06-JUN-1997; US-048916.
PR 06-JUN-1997; US-048962.
PR 06-JUN-1997; US-048970.
PR 06-JUN-1997; US-048974.
PR 06-JUN-1997; US-049373.
PR 05-SEP-1997; US-057584.
PR 05-SEP-1997; US-057629.
PR 05-SEP-1997; US-057642.
PR 05-SEP-1997; US-057645.
PR 05-SEP-1997; US-057648.
PR 05-SEP-1997; US-057651.
PR 05-SEP-1997; US-057662.
PR 05-SEP-1997; US-057668.
PR 05-SEP-1997; US-057762.
PR 05-SEP-1997; US-057765.
PR 05-SEP-1997; US-057771.
PR 05-SEP-1997; US-057776.
PR 06-JUN-1997; US-048876.
PR 06-JUN-1997; US-048880.
PR 06-JUN-1997; US-048883.
PR 06-JUN-1997; US-048892.
PR 06-JUN-1997; US-048895.
PR 06-JUN-1997; US-048898.
PR 06-JUN-1997; US-048901.
PR 06-JUN-1997; US-048917.
PR 06-JUN-1997; US-048963.
PR 06-JUN-1997; US-048971.
PR 06-JUN-1997; US-049019.
PR 06-JUN-1997; US-049374.
PR 05-SEP-1997; US-057627.
PR 05-SEP-1997; US-057634.
PR 05-SEP-1997; US-057643.
PR 05-SEP-1997; US-057646.
PR 05-SEP-1997; US-057649.
PR 05-SEP-1997; US-057654.
PR 05-SEP-1997; US-057666.
PR 05-SEP-1997; US-057760.
PR 05-SEP-1997; US-057763.
PR 05-SEP-1997; US-057769.
PR 05-SEP-1997; US-057774.
PR 05-SEP-1997; US-057777.
PR (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR WPI; 99-059865/05.
DR P-PSDB; W88616.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 4; Page 345; 772pp; English.
CC The invention relates to nucleic acid sequences (V84411 to V84633)
CC encoding human secreted proteins (W88534 to W88756). The secreted protein
CC gene sequences are deposited with the ATCC under deposit numbers ATCC

```

---

```

CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung
CC or thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The present sequence represents a gene encoding a human secreted protein
CC (see descriptor line for gene number and clone identification).
CC Sequence 744 BP; 172 A; 174 C; 217 G; 175 T;
SQ

Query Match 72.0%; Score 14.4; DB 1; Length 744;
Best Local Similarity 93.8%; Pred. No. 87;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 cttccctaaaggagg 20
   ||||| |||||
Db 124 CTCCCTGAGGAGG 139

RESULT 8
N80975
ID N80975 standard; DNA; 780 BP.
AC N80975;
DT 12-SEP-1990 (first entry)
DE Probe PCS.7 for screening for cystic fibrosis-associated RFLP in HhaI
DE digests of human chromosome 7q 22-31 DNA
KW Probe PCS.7; cystic fibrosis screening; human chromosome 7q 22-31;
KW restriction fragment length polymorphism (RFLP); cosmid CNX.4; ds.
OS Homo sapiens.
PH Key Location/Qualifiers
FT misc_feature 213
FT /*tag= a
FT /note="polymorphism occurs here, see CC"
FT
FT EP-288299-A.
FT 26-OCT-1988.
FT 22-APR-1988; 303645
FT 23-APR-1987; GB-009652.
FT (STMA) St Marys Hospital.
FT Williamson R;
FT WPI; 88-301321/43.
FT Region of human genomic DNA from chromosome 7722-31 -
FT used for producing a DNA hybridisation probe for cystic fibrosis
FT risk analysis and diagnosis
FT Disclosure; Page 12; 15pp; English.
CC Probe PCS.7 is from a 2.0 kb NotI/HindIII fragment derived from cosmid
CC CNX.4, which contains a putative HpaII tiny fragments (HTF) island. It
CC detects a frequent RFLP (Al 0.56, A2 0.44) in HhaI digests. In 70 out of
CC 71 chromosomes studied, the CF mutation is associated with the PCS.7 A2
CC allele (0.47Kb). The polymorphism occurs at tag a (see FT) where the
CC base pair GC provides HhaI site, while the alternative of AT eliminates
CC the site. The probe, which is claimed, can be used in conjunction with
CC haplotyping or carrier exclusion and would enable about two-thirds of the
CC population to be excluded from significant risk.
CC Sequence 780 BP; 125 A; 261 C; 230 G; 164 T;
SQ

Query Match 72.0%; Score 14.4; DB 1; Length 780;
Best Local Similarity 93.8%; Pred. No. 87;

```

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggaattccctaaagg 16  
 ||| ||||| |||||

Db 246 GGAGCTTCCTAAGG 261

# RESULT 9

T68000  
 ID T68000 standard; DNA; 1041 BP.

AC T68000.

DT 16-JUL-1997 (first entry)

DE H. pylori inner membrane protein ORF 06ep10306orf3.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacterium; life cycle; activator;

KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;

KW diagnosis; ds.

OS Helicobacter pylori.

FH key Location/Qualifiers

FT cds 1..1041

FT /\*tag= a

FT /note= "no stop codon given"

PN WO9640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR ) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI; 97-052306/05.

DR P-PSDB; W20747.

PT Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

PS Claim 1; Page 838; 1481pp; English.

CC The present sequence encodes a H. pylori inner membrane protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To

CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.

CC Sequence 1041 BP; 312 A; 200 C; 212 G; 317 T;

QY 4 acttcctctaaaggag 19

||||| ||||| |||||

Db 46 ACTTGCTTAAGGAG 61

# RESULT 10

V25046

ID V25046 standard; DNA; 1146 BP.

AC V25046;

DT 03-JUL-1998 (first entry)

DE H. pylori cytoplasmic protein ORF 03ael0804\_21698400\_c2\_32.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacteria; life cycle; activator;

KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.

OS Helicobacter pylori.

FH key Location/Qualifiers

FT cds 1..1146

FT WO9737044-A1.  
 FT 09-OCT-1997.  
 PD 27-MAR-1997; U05223.  
 PF 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 PA (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI; 97-503122/46.  
 DR P-PSDB; W55637.  
 DT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection

PS Claims 5,6,42; Page 412; 1145pp; English.

CC This sequence encodes a H. pylori cytoplasmic protein. The protein may be

CC used in a vaccine to prevent or treat H. pylori infection or to identify

CC H. pylori polypeptide binding compounds, useful as potential H. pylori

CC life cycle activators or inhibitors. The DNA and probes derived from it

CC may be used for the identification of H. pylori in a sample and the

CC diagnosis of H. pylori infection. Nucleic acid sequences complementary to

CC the DNA act as antisense sequences and can be used to prevent the

CC translation of H. pylori mRNA. Antibodies against the protein can be used

CC in immunoassays to evaluate the abundance and distribution of

CC H. pylori-specific antigens. The genomic sequence of H. pylori

CC (ATCC 55679) was determined from overlapping contigs generated by

CC mechanically shearing the bacterial DNA. The sequences were analysed

CC for ORF of at least 180 nucleotides, and the predicted coding regions

CC defined by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or exported

CC membrane proteins. Having identified and determined the sequences of

CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in E. coli

CC hosts.

CC Sequence 1146 BP; 402 A; 201 C; 216 G; 327 T;

QY 4 acttcctctaaaggag 19

||||| ||||| |||||

Db 1067 ACTTGCTTAAGGAG 1082

Query Match 72.0%; Score 14.4; DB 1; Length 1146;

Best Local Similarity 93.8%; Pred. No. 90;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

# RESULT 11

V24720

ID V24720 standard; DNA; 1158 BP.

AC V24720;

DT 15-JUN-1998 (first entry)

DE H. pylori ORF 06ep10306orf11.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacteria; life cycle; activator;

KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.

OS Helicobacter pylori.

FH key Location/Qualifiers

FT cds 1..1158

FT /\*tag= a

FT /note= "no stop codon given"

PN WO9737044-A1.

PD 09-OCT-1997.

PF 27-MAR-1997; U05223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PA (ASTR ) ASTRA AB.

PI Alm RA, Smith D;  
 DR WPI: 97-503122/46.  
 DR P-PSDB: W55311.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claims 5,6; Page 203; 1145pp; English.  
 CC This sequence encodes a H. pylori protein of unspecified function.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 SQ Sequence 1158 BP; 406 A; 204 C; 219 G; 329 T;  
 Query Match 72.0%; Score 14.4; DB 1; Length 1158;  
 Best Local Similarity 93.8%; Pred. No. 90;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 4 attccctctaaaggag 19  
 Db 1082 ACTGCTCTAAAGGAG 1097  
 RESULT 12  
 X14148  
 ID X14148 standard; DNA; 1350 BP.  
 AC X14148;  
 DT 31-MAR-1999 (first entry)  
 DE H. pylori GHP0 422 gene.  
 KW GHP0 protein; Helicobacter Infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease; ss.  
 OS Helicobacter pylori.  
 FH Key Location/Qualifiers  
 FT CDS 60..1205  
 FT /\*tag= a  
 FT W09843478-AL.  
 PN 08-OCT-1998.  
 PD 08-OCT-1998.  
 PF 01-APR-1998; U06371.  
 PR 29-JUL-1997; US-902615.  
 PR 01-APR-1997; US-833457.  
 PR 24-JUN-1997; US-881227.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
 DR WPI: 98-542293/46.  
 DR P-PSDB: W98429.  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 PS Claim 1; Page 724-726; 2054pp; English.  
 CC This sequence represents a polynucleotide of the invention. It was  
 CC isolated from Helicobacter pylori and encodes a H.pylori GHP0 protein.  
 CC The polypeptides can be used for preventing or treating Helicobacter  
 CC infections, and gastroduodenal diseases associated with these  
 CC infections, including acute, chronic, and atrophic gastritis, and peptic  
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used  
 CC for the production of antibodies. The products can also be used for

CC detection and diagnosis.  
 SQ Sequence 1350 BP; 464 A; 223 C; 257 G; 406 T;  
 Query Match 72.0%; Score 14.4; DB 1; Length 1350;  
 Best Local Similarity 93.8%; Pred. No. 91;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 4 attccctctaaaggag 19  
 Db 1126 ACTGCTCTAAAGGAG 1141  
 RESULT 13  
 Q43719  
 ID Q43719 standard; DNA; 1404 BP.  
 AC Q43719;  
 DT 25-SEP-1993 (first entry)  
 DE Sequence which hybridises specifically to bovine male DNA.  
 KW Bovine; embryo; sex selection; ss.  
 OS Bos primigenius, strain Holstein.  
 PN EP-546762-A.  
 PD 16-JUN-1993.  
 PF 02-DEC-1992; 311013.  
 PR 13-DEC-1991; JP-352032.  
 PR (TOH-) ITOHAM FOODS INC.  
 PA Itagaki Y, Kudo T, Nakamura T, Sato S, Suto S;  
 PI WPI: 93-190118/24.  
 DR Sexing bovine embryos - by amplifying specific DNA sequences by  
 PT polymerase chain reaction using male-specific and gender-neutral  
 PT primers  
 PS Claim 1; Pages 9-10; 40pp; English.  
 CC Q43719 is the DNA sequence of a plasmid which hybridises  
 CC specifically to bovine male genomic (BMG) DNA. The plasmid having  
 CC this DNA sequence was transfected into E. coli and the transfectant  
 CC E.c.118-bms1 has been deposited as FERM BP-4095. Q43719 was used as  
 CC a probe to screen a bovine male genomic library. The plaque  
 CC hybridization gave 28 positive clones. One of them was picked up and  
 CC its DNA was extracted and EcoRI digested into 2 fragments (Q43720,  
 CC Q43721). Both DNAs specifically hybridize to the male DNA. They were  
 CC introduced into E. coli and the resultant recombinants E.c.gem-bms1  
 CC and E.c.gem-bms2 have been deposited as FERM BP-4089 and FERM BP-4090  
 CC respectively. One clone was selected from the 28 clones (see  
 CC Q43722). This DNA sequence hybridized to both male and female DNA,  
 CC but images were different from each other. DNA having sequence  
 CC Q43722 was introduced into E. coli and the recombinant E.c.gem-bms3  
 CC has been deposited as FERM BP-4091. When male-specific clones were  
 CC explored 20 clones were found to hybridize to both male and female  
 CC DNA in the Southern blot analysis. Three clones which were expected  
 CC to be repetitious were selected and used as probes for the Southern  
 CC blot analyses (see Q43723, Q43724, Q43725). DNA consisting of these  
 CC sequences were introduced into E. coli and the resultant  
 CC recombinants, E.c.118-bmfl, E.c.118-bmf2, E.c.118-bmf3, have been  
 CC deposited as FERM BP-4092, FERM BP-4093 and FERM BP-4094.  
 SQ Sequence 1404 BP; 385 A; 289 C; 310 G; 420 T;  
 Query Match 72.0%; Score 14.4; DB 1; Length 1404;  
 Best Local Similarity 93.8%; Pred. No. 91;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 3 aactccctctaaaggag 18  
 Db 381 AACTCTCCCAAAAGGGA 396  
 RESULT 14  
 X16664  
 ID X16664 standard; DNA; 1404 BP.  
 AC X16664;  
 DT 30-APR-1999 (first entry)  
 DE Bovine Y chromosome-specific DNA #1.  
 DE Bovine; Bos taurus; Holstein; Y chromosome; male; female; ds.

OS Bos taurus.  
PN J08256778-A.  
PD 08-OCT-1996.  
PF 23-MAR-1995; 088596.  
PR 23-MAR-1995; JP-088596.  
PA (KACH-) KACHIKU JUSEIRAN ISHOKU GIJUTSU KENKYU.  
DR WPI: 96-500367/50.  
PT Bovine Y chromosome-specific DNA - used as a marker for the DNA  
common to two sexes  
PS Claim 1; Page 9; 19pp; Japanese.  
CC The present invention describes a marker containing a DNA having at  
least one sequence of X16664 to X16670 or its portion. Also described  
is a method for identifying or specifying a biosample by using the  
above marker, and a method for the DNA present the 5' and/or the 3',  
side adjacent to at least one sequence of X16664 to X16670 by using a  
DNA having part of the sequence as the primer. It can specify or  
identify a biosample. For example, it can identify a sample to be  
bovine-derived. It can specify nucleus and cytoplasm, sex chromosome  
and autosome, X chromosome and Y chromosome, short arm and long arm.  
CC In an example, bovine genomic DNA was isolated from male and female  
bovine liver. Sheep, swine and caprine DNA were isolated. DNA of  
Seq. Nos. 1-7 were isolated and purified. The bovine male specificity of  
CC male bovine-derived DNA was examined.  
SQ Sequence 1404 BP; 385 A; 290 C; 309 G; 420 T;

Query Match 72.0%; Score 14.4; DB 1; Length 1404;  
Best Local Similarity 93.8%; Pred. No. 91;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aacttcctctaaaggga 18  
||||||| |||||  
Db 381 AACTTCCCAAAAGGGA 396

RESULT 15

Q98950  
ID Q98950 standard; DNA; 3134 BP.  
AC Q98950;  
DT 25-APR-1996 (first entry)  
DE bg8 insert encoding a B-G subregion fragment of fowl MHC.  
KW Domesticated fowl; chicken; turkey; pheasants; B-G antigen; MHC;  
KW major histocompatibility complex; haplotyping; Marek's disease;  
restriction fragment length polymorphism; bg8 insert; probe; ds.  
OS Synthetic.  
PN US5451670-A.  
PD 19-SEP-1995.  
PF 30-JUN-1987; 068176.  
PR 30-JUN-1987; US-068176.  
PR 09-DEC-1987; US-130529.  
PR 23-JUN-1988; US-210405.  
PR 28-SEP-1989; US-413301.  
PR 27-SEP-1990; US-588922.  
PR 22-APR-1991; US-688326.  
PR 07-APR-1992; US-865662.  
PA (CITY ) CITY OF HOPE.  
PI Miller MM.  
DR WPI: 95-336319/43.  
PT New DNA encoding the B-G antigen of fowl MHC - and derived probes,  
useful for haplotyping and determining genotype at specific loci  
PS Claim 5; Columns 23-28; 61pp; English.  
CC The DNA sequences Q98944-55 contain a polymorphic portion of the  
coding region of domestic fowl's (i.e. chicken, turkey, pheasant)  
CC MHC B-G subregion (Q98944 encodes the B-G subregion fragment  
R89581). The sequences are used as probes to detect restriction  
CC fragment length polymorphism patterns typical of various B-G  
CC alleles, e.g. resistance to Marek's (and other) disease, general  
CC fitness and productivity, all are related to MHC haplotype. This  
CC haplotyping method has the advantage of not requiring alloantisera.  
SQ Sequence 3134 BP; 818 A; 718 C; 777 G; 821 T;

Query Match 72.0%; Score 14.4; DB 1; Length 3134;

Best Local Similarity 93.8%; Pred. No. 97;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 5 cttccctaaaggagg 20  
||||||| |||||  
Db 1110 CTTCCCTAAAGCGAGG 1125

Search completed: May 23, 2000, 09:42:13  
Job time: 8845 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:21:43 ; Search time 1428.99 Seconds  
(without alignments)  
56.729 Million cell updates/sec

Title: US-08-945-805-3  
Perfect score: 20  
Sequence: 1 ggaacttcctaaaggagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :	
EST:*	
1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*
44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	gb_est33:*
60:	gb_est34:*
61:	gb_est35:*
62:	gb_est36:*
63:	gb_est37:*
64:	gb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	gb_est39:*
70:	gb_est40:*
71:	gb_est41:*
72:	gb_est42:*
73:	gb_est43:*
74:	gb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	gb_est45:*
80:	gb_est46:*
81:	gb_est47:*
82:	gb_gss1:*
83:	gb_gss2:*
84:	gb_gss3:*
85:	gb_gss4:*
86:	em_gss1:*
87:	em_gss2:*
88:	em_gss3:*
89:	em_gss4:*
90:	gb_gss5:*
91:	gb_gss6:*
92:	gb_gss7:*
93:	gb_gss8:*
94:	gb_gss9:*
95:	em_gss5:*
96:	em_gss6:*
97:	em_gss7:*
98:	em_gss8:*
99:	em_gss9:*
100:	em_gss10:*
101:	em_gss11:*
102:	gb_gss10:*
103:	gb_gss11:*
104:	em_gss12:*
105:	gb_gss12:*
106:	gb_gss13:*
107:	gb_gss14:*
108:	gb_gss15:*
109:	gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query
--------	---	-------

No.	Score	Match	Length	DB	ID	Description
c 1	18	90.0	475	49	AV006799	AV006799 AV006799
c 2	17.4	87.0	176	47	AI506735	AI506735 vn76a08.x
c 3	17.4	87.0	425	41	AI035972	AI035972 ub50c11.r
c 4	16.8	84.0	411	61	AI874221	AI874221 wm50d11.x
c 5	16.4	82.0	236	70	AV246642	AV246642 AV246642
c 6	16.4	82.0	313	20	T06661	T06661 EST04550 Fe
c 7	16.4	82.0	456	108	AQ567946	AQ567946 HS_2135_B
c 8	16.4	82.0	850	83	AF101883	AF101883 Homo sapi
c 9	15.8	79.0	242	73	AV367114	AV367114 AV367114
c 10	15.8	79.0	313	60	AV166828	AV166828 AV166828
c 11	15.8	79.0	318	79	AW298554	AW298554 UI-H-BW0-
c 12	15.8	79.0	354	81	AW400824	AW400824 Lamd1Sest
c 13	15.8	79.0	397	109	AQ596947	AQ596947 HS_5207_B
c 14	15.8	79.0	433	46	AI447647	AI447647 mq76e11.x
c 15	15.8	79.0	444	84	BI5892	BI5892 34666.tv CI
c 16	15.8	79.0	470	90	AQ812869	AQ812869 HS_5291_A
c 17	15.8	79.0	512	84	B59255	B59255 CIT-HSP-201
c 18	15.8	79.0	530	38	AA790378	AA790378 vv05h07.r
c 19	15.8	79.0	573	81	AW399458	AW399458 EST309958
c 20	15.8	79.0	608	82	FR0031079	FR0031079 Fugu rubr
c 21	15.8	79.0	1061	91	AQ900282	AQ900282 HS_3168_A
c 22	15.4	77.0	246	71	AV273782	AV273782 AV273782
c 23	15.4	77.0	355	33	AA384199	AA384199 EST97728
c 24	15.4	77.0	363	34	AA528363	AA528363 ne83f02.s
c 25	15.4	77.0	391	39	AA836485	AA836485 od38d10.s
c 26	15.4	77.0	396	22	R38287	R38287 yc93e10.sl
c 27	15.4	77.0	420	84	AA52890	AA52890 zx36g12.s
c 28	15.4	77.0	420	84	B36069	B36069 HS-1031-B1-
c 29	15.4	77.0	435	33	AA442006	AA442006 zw56a03.r
c 30	15.4	77.0	446	25	N63771	N63771 za19e10.sl
c 31	15.4	77.0	494	85	AQ731200	AQ731200 HS_5531_A
c 32	15.4	77.0	529	51	AI743902	AI743902 wg54c11.x
c 33	15.4	77.0	554	59	AI768219	AI768219 wg82f01.x
c 34	15.4	77.0	568	91	AQ890836	AQ890836 HS_3063_A
c 35	15.4	77.0	574	64	AW057533	AW057533 wy60d01.x
c 36	15.4	77.0	575	81	AW450328	AW450328 UI-H-B13-
c 37	15.4	77.0	609	82	AG009571	AG009571 Homo sapi
c 38	15.4	77.0	644	59	AI775826	AI775826 EST256926
c 39	15.4	77.0	654	109	AQ621882	AQ621882 HS_3107_B
c 40	15.4	77.0	666	64	AL039412	AL039412 DKFZp434F
c 41	15.4	77.0	735	61	AL040178	AL040178 DKFZp434F
c 42	15.2	76.0	145	24	H91587	H91587 yw29f02.r1
c 43	15.2	76.0	183	51	AV064089	AV064089 AV064089
c 44	15.2	76.0	209	73	AV375710	AV375710 AV375710
c 45	15.2	76.0	215	59	AV080222	AV080222 AV080222

## ALIGNMENTS

RESULT 1  
AV006799/c  
LOCUS AV006799 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA  
DEFINITION clone 1110001K06, mRNA sequence.  
ACCESSION AV006799  
VERSION AV006799.1 GI:4783786  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 475)  
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Tateno,M., Tomaru,Y., Tomisaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

TITLE  
JOURNAL  
COMMENT

RIKEN Mouse ESTs  
Unpublished (1999)  
On May 7, 1998 this sequence version replaced gi:3121513.  
Contact: Chie Owa  
Genome Science Laboratory  
RIKEN  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098  
Email: genome-res@rtrc.riken.go.jp  
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtrc.riken.go.jp>) for further details.

FEATURES  
source

Location/Qualifiers  
1. 475  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="1110001K06"  
/sex="mixed"  
/dev\_stage="18-day embryo"  
150 a 89 c 111 g 113 t 12 others

BASE COUNT  
ORIGIN

Query Match 90.0%; Score 18; DB 49; Length 475;  
Best Local Similarity 94.7%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 gaacttcctaaaggagg 20  
Db 145 GAACCTCCNAAGGGAGG 127

## RESULT 2

AI506735  
LOCUS AI506735 176 bp mRNA EST 11-MAR-1999  
DEFINITION vn76a08.x1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA  
clone IMAGE:1037846 3', mRNA sequence.  
ACCESSION AI506735  
VERSION AI506735.1 GI:4404586  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 176)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swallier,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
On May 13, 1998 this sequence version replaced gi:3137568.  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:577870  
This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
Seq primer: primer name ambiguous  
High quality sequence stop: 62.

FEATURES  
source

Location/Qualifiers  
1. 176  
/organism="Mus musculus"  
/strain="B6D2 F1/3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1037846"  
/clone\_lib="Knowles Solter mouse blastocyst B1"  
/tissue\_type="blastocyst"  
/dev\_stage="embryo (pre-implantation)"  
/lab\_host="DH10B"  
/note="Organ: embryo; Vector: pSPORT; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(dt): 5'-CGGTCGACGCGACGCTTTT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies). Two different size selections: B1 (larger inserts) and B3."

BASE COUNT 44 a 57 c 41 g 34 t  
ORIGIN

Query Match 87.0%; Score 17.4; DB 47; Length 176;  
Best Local Similarity 94.7%; Pred. No. 28;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaacttcctaaaggagg 20

||||||| |||||||

Db 6 GAATTCCTCCAAAGGAGG 24

## RESULT 3

AI035972/c

LOCUS

ub50c11.r1 Soares\_mammary\_gland\_NbMMG Mus musculus cDNA clone

IMAGE:1381172 5', mRNA sequence.

AI035972

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 425)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Fan,F., Underwood,K., Moore,B.,

Thelising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

Waterston,R.

Unpublished (1996)

On Sep 12, 1996 this sequence version replaced gi:1400806.

Contact: Marra M/Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mousest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:903640

Seq primer: -28ml3 rev2 ET from Amersham.

Location/Qualifiers

1. 425

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:1381172"

/clone\_lib="Soares\_mammary\_gland\_NbMMG"

/sex="male"

/tissue\_type="mammary gland"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTGGAGCGCCGCCGAATGCTTTT-3'] and cloned into T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 94 a 103 c 125 g 103 t  
ORIGIN

Query Match 87.0%; Score 17.4; DB 41; Length 425;  
Best Local Similarity 94.7%; Pred. No. 32;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gaacttcctaaaggagg 20

||||||| |||||||

Db 302 GAATTCCTCCAAAGGAGG 284

## RESULT 4

AI874221

LOCUS

wm50d11.x1 NCI\_CGAP\_Ut4 Homo sapiens cDNA clone IMAGE:2439381 3'

similar to gb:X01677 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE,

LIVER (HUMAN);, mRNA sequence.

AI874221

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 411)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On May 18, 1998 this sequence version replaced gi:3137087.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 322.

Location/Qualifiers

1. 411

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2439381"

/clone\_lib="NCI\_CGAP\_Ut4"

/tissue\_type="serous papillary carcinoma, high grade, 2

pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.48 kb. Life Technologies catalog #:

11542-016"

BASE COUNT 82 a 87 c 138 g 104 t  
ORIGIN

```

Query Match      84.0%; Score 16.8; DB 61; Length 411;
Best Local Similarity 90.0%; Pred. No. 66;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggagg 20
    ||||| ||| |||||
Db 144 GGAACGCCCAAAAGGGAGG 163

RESULT 5
AV246642
LOCUS AV246642 RIKEN full-length enriched, 0 day neonate head Mus EST 04-NOV-1999
DEFINITION musculus cDNA clone 4832413L08 3', mRNA sequence.
ACCESSION AV246642
VERSION AV246642.1 GI:6234101
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 236)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
Taunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (1999)
On Dec 20, 1995 this sequence version replaced gi:1133307.
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kusunagi,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source Location/Qualifiers
1..236
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4832413L08"
/clone_lib="RIKEN full-length enriched, 0 day neonate
head"
/sex="mixed"
/tissue_type="head"
/dev_stage="0 day neonate"
/lab_host="DH10B"

Query Match      82.0%; Score 16.4; DB 70; Length 236;
Best Local Similarity 94.4%; Pred. No. 99;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aacttcctaaaggagg 20
    ||||| |||||
Db 63 AACCTCGTAAGGGAGG 80

RESULT 6
T06661/c
LOCUS T06661 313 bp mRNA EST 30-JUN-1993
DEFINITION EST04550 Fetal brain, Stragogene (cat#936206) Homo sapiens cDNA
clone HFBDW51, mRNA sequence.
ACCESSION T06661
VERSION T06661.1 GI:317810
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 313)
Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
3,400 expressed sequence tags identify diversity of transcripts
from human brain
Nature Genet. 4, 256-267 (1993)
JOURNAL MEDLINE
COMMENT Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.
FEATURES
source Location/Qualifiers
1..313
/organism="Homo sapiens"
/db_xref="ATCC (inhost):83315"
/db_xref="taxon:9606"
/clone="HFBDW51"
/clone_lib="Fetal brain, Stragogene (cat#936206)"
/notes="Vector: LambdaZAP-II; 17-18 wk gestation, female;
oligo-dT + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average inser size."

BASE COUNT 78 a 64 c 49 g 120 t
ORIGIN

Query Match      82.0%; Score 16.4; DB 20; Length 313;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

BASE COUNT 64 a 38 c 51 g 83 t
ORIGIN

```

/note="Site.1: Sali; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATAATAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1."

```

Qy 3 aacttcctaaaggagggg 20
I | ||||| ||||| |||||
Db 91 AGCTTCCTCAAGGAGG 74

RESULT 7
AQS67946      456 bp   DNA       GSS          01-JUN-1999
LOCUS         HS_2135_B1_H06_T7C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION    sapiens genomic clone Plate=2135 Col=11 Row=P, genomic survey
sequence.
ACCESSION     AQ567946
VERSION       AQ567946.1 GI:4961309
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE     1 (bases 1 to 456)
AUTHORS       Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
               Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
               Hood,L.
TITLE         Sequence-tagged connectors: A sequence approach to mapping and
               scanning the human genome
JOURNAL       Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE       99380589
COMMENT       On Sep 10, 1998 this sequence version replaced gi:3552895.
               Contact: Mahairas GG, Wallace JC, Hood L
               High Throughput Sequencing Center
               University of Washington
               401 Queen Anne Avenue North, Seattle, WA 98109, USA
               Tel: (206) 616-3618
               Fax: (206) 616-3887
               Email: jwallace@u.washington.edu
               Clones may be purchased from Research Genetics (info@resgen.com).
               BAC end Web Server: http://www.htsc.washington.edu
               Plate: 2135 row: P column: 11
               Seq primer: T7
               Class: BAC ends
               High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source           1..456
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_plate="2135 Col=11 Row=P"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBelBAC11; BAC Clones In
BASE COUNT         105 a 132 c 89 g 130 t
ORIGIN              E-Coli DH10B"

Query Match      82.0%; Score 16.4; DB 108; Length 456;
Best Local Similarity 94.4%; Pred. No. 1.le+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggaactccctaaaggga 18
I | ||||| ||||| |||||
Db 88 GGAATTCCTCAAGGA 105

RESULT 8
AF101883      850 bp   DNA       GSS          08-MAR-1999
LOCUS         AF101883
DEFINITION    Homo sapiens chromosome 11 clone pTWB9n12.28SP6 map llp15.5,
sequence.
ACCESSION     AF101883
VERSION       AF101883.1 GI:4193722
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE     1 (bases 1 to 850)
AUTHORS       Bepler,G., O'Briant,K.C., Kim,Y.-C., Schreiber,G. and Pitterle,D.M.
               A 1.4-Mb high-resolution physical map and contig of chromosome
               segment llp15.5 and genes in the LOHLIA metastasis suppressor
               region
JOURNAL       Genomics 55 (2), 164-175 (1999)
MEDLINE       99134294
DEFINITION    2 (bases 1 to 850)
AUTHORS       Bepler,G., O'Briant,K.C., Kim,Y.-C., Schreiber,G. and Pitterle,D.M.
               Direct Submission
TITLE         Submitted (27-OCT-1998) Medicine and Radiology, Duke University
               Medical Center, Box 2610, MSRB, Room 117, Durham, NC 27710, USA
               Location/Qualifiers
FEATURES             Location/Qualifiers
     source           1..850
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="llp15.5"
                     /clone="pTWB9n12.28SP6"
                     /note="part of a 1.4 megabase contig including the LOHLIA
metastasis suppressor region Bin M"
BASE COUNT         225 a 157 c 218 g 240 t 10 others
ORIGIN

Query Match      82.0%; Score 16.4; DB 83; Length 850;
Best Local Similarity 94.4%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggaactccctaaaggga 18
I | ||||| ||||| |||||
Db 444 GGAATTCCTCAAGGA 461

RESULT 9
AV367114      242 bp   mRNA          EST          14-NOV-1999
LOCUS         AV367114 RIKEN full-length enriched, 16 days embryo lung Mus
DEFINITION    musculus cDNA clone 8430427L19 3', mRNA sequence.
ACCESSION     AV367114
VERSION       AV367114.1 GI:6414761
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE     1 (bases 1 to 242)
AUTHORS       Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
               Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Horii,F.,
               Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
               Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
               Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
               Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
               Shibata,Y., Shigenoto,Y., Shiraki,T., Sogabe,Y., Sugahara,N.,
               Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N.,
               Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
               Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
               RIKEN Mouse ESTs (Konno.H., et al.)
               Unpublished (1999)
               On Jul 8, 1999 this sequence version replaced gi:5422462.
               Contact: Yoshihide Hayashizaki
               Genome Exploration Research Group, Life Science Tsukuba Center,
               Genome Science Laboratory
               The Institute of Physical and Chemical Research (RIKEN), Genomic
               Sciences Center
               3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
               Tel.: +81-298-36-9013
               Fax: +81-298-36-9098
               Email: genome-resctr@riken.go.jp,
               URL:http://genome.rtc.riken.go.jp/
               Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
               Yoneda,Y.

```

[illegible]



```

FEATURES
  Location/Qualifiers
    1..318
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
BASE COUNT      75 a      63 g      106 t
ORIGIN
      1..318
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
Query Match      79.0%; Score 15.8; DB 79; Length 318;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 gaacttcctaaaggagg 20
      |||||||
Db      282 GAAC TTCCTACGGGAGG 300

RESULT 12
AW400824      354 bp      mRNA      EST      07-FEB-2000
LOCUS      Lamd1Sest410est L.digitata sporophyte Lambda ZapII Laminaria
DEFINITION      digitata cDNA similar to ribosomal protein L36, mRNA sequence.
ACCESSION      AW400824
VERSION
KEYWORDS
SOURCE
ORGANISM
  Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group;
  Laminaria digitata.
  Laminaria digitata
  Eukaryota; stramenopiles; Laminariales; Laminariaceae; Laminaria.
REFERENCE
  1 (bases 1 to 354)
  Crepiniau,F., Roscoe,T., Kaas,R., Kloareg,B. and Boyen,C.
  Characterisation of complementary DNAs from the Expressed Sequence
  Tag analysis of life cycle stages of Laminaria digitata
  (Phaeophyceae)
  Unpublished (2000)

JOURNAL
  Location/Qualifiers
    1..318
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
BASE COUNT      75 a      63 g      106 t
ORIGIN
      1..318
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
Query Match      79.0%; Score 15.8; DB 81; Length 354;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 gaacttcctaaaggagg 20
      |||||||
Db      26 GAACGTCCTCAAGAGAGG 44

RESULT 13
AQ596947/c      397 bp      DNA      GSS      08-JUN-1999
LOCUS      HS_5207.B2.B02.SP6E.RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION      genomic clone Plate-783 Col-4 Row-D, genomic survey sequence.
ACCESSION      AQ596947
VERSION
KEYWORDS
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  99380589
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
  or from Resear h Genetics (info@resgen.com). BAC end Web Server:
  http://www.htsc.washington.edu
  Plate: 783 row: D column: 4
  Seq primer: SP6
  Class: BAC ends
  High quality sequence stop: 397.
  Location/Qualifiers
    1..397
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
BASE COUNT      75 a      63 g      106 t
ORIGIN
      1..318
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
Query Match      79.0%; Score 15.8; DB 81; Length 354;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 gaacttcctaaaggagg 20
      |||||||
Db      26 GAACGTCCTCAAGAGAGG 44

RESULT 13
AQ596947      397 bp      DNA      GSS      08-JUN-1999
LOCUS      HS_5207.B2.B02.SP6E.RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION      genomic clone Plate-783 Col-4 Row-D, genomic survey sequence.
ACCESSION      AQ596947
VERSION
KEYWORDS
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  99380589
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
  or from Resear h Genetics (info@resgen.com). BAC end Web Server:
  http://www.htsc.washington.edu
  Plate: 783 row: D column: 4
  Seq primer: SP6
  Class: BAC ends
  High quality sequence stop: 397.
  Location/Qualifiers
    1..397
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
BASE COUNT      75 a      63 g      106 t
ORIGIN
      1..318
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
Query Match      79.0%; Score 15.8; DB 81; Length 354;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 gaacttcctaaaggagg 20
      |||||||
Db      26 GAACGTCCTCAAGAGAGG 44

RESULT 13
AQ596947      397 bp      DNA      GSS      08-JUN-1999
LOCUS      HS_5207.B2.B02.SP6E.RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION      genomic clone Plate-783 Col-4 Row-D, genomic survey sequence.
ACCESSION      AQ596947
VERSION
KEYWORDS
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  99380589
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
  or from Resear h Genetics (info@resgen.com). BAC end Web Server:
  http://www.htsc.washington.edu
  Plate: 783 row: D column: 4
  Seq primer: SP6
  Class: BAC ends
  High quality sequence stop: 397.
  Location/Qualifiers
    1..397
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 L1AM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (50% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of
the driver population). Subtraction was performed as
previously described (Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_L1B=NCI_CGAP_Brn50
TAG_TISSUE=Brain
TAG_SEQ=TTTTCG"
BASE COUNT      75 a      63 g      106 t
ORIGIN
      1..318
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2732370"
      /clone_lib="NCI_CGAP_Sub6"
      /lab_host="DH10B (Life Technologies)"
      /note="Vector: pT730-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI_CGAP_Brn50,
NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1. The NCI_CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool
(NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
pool 1 L1AM 3164-3167, 3716-3720, 3733-373
```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:40:27 ; Search time 71.7 Seconds  
(without alignments)  
36.258 Million cell updates/sec

Title: US-08-945-805-3  
Perfect score: 20  
Sequence: 1 ggaacttcctaaaggagg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2.6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2.6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2.6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2.6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.4	77.0	3559	US-08-211-312-1	Sequence 1, Appl
C 2	15.4	77.0	3559	US-08-472-285-1	Sequence 1, Appl
C 3	14.8	74.0	24979	US-08-147-777-3	Sequence 3, Appl
C 4	14.8	74.0	24979	PCT-US93-03985-3	Sequence 3, Appl
C 5	14.4	72.0	1404	US-07-984-044A-1	Sequence 1, Appl
C 6	14.4	72.0	1404	US-08-458-393-1	Sequence 1, Appl
C 7	14.4	72.0	3134	US-07-865-662F-7	Sequence 7, Appl
C 8	14.4	72.0	4488	US-08-441-430-1	Sequence 1, Appl
C 9	14.4	72.0	5599	US-08-477-451-9	Sequence 9, Appl
C 10	14.4	72.0	5599	US-08-477-451-13	Sequence 13, Appl
C 11	14.4	72.0	19932	US-08-477-451-25	Sequence 25, Appl
C 12	14.2	71.0	1668	US-08-901-547A-1	Sequence 1, Appl
C 13	14.2	71.0	5339	US-09-073-354-6	Sequence 6, Appl
C 14	14.2	71.0	5339	US-09-073-354-6	Sequence 6, Appl
C 15	14.2	71.0	5342	US-08-902-632-1	Sequence 1, Appl
C 16	14.2	71.0	5342	US-08-902-632-1	Sequence 1, Appl
C 17	14.2	71.0	5342	US-09-073-354-5	Sequence 5, Appl
C 18	14.2	71.0	5342	US-09-073-354-5	Sequence 5, Appl
C 19	14.2	71.0	7210	US-08-257-963B-10	Sequence 10, Appl
C 20	14.2	71.0	7210	PCT-US95-07201-10	Sequence 10, Appl
C 21	14.2	71.0	22481	PCT-US95-07201-43	Sequence 43, Appl
C 22	13.8	69.0	453	US-08-967-101-123	Sequence 123, App
C 23	13.8	69.0	453	US-08-592-541-123	Sequence 123, App
C 24	13.8	69.0	1003	US-08-967-101-151	Sequence 151, App
C 25	13.8	69.0	1003	US-08-592-541-151	Sequence 151, App
C 26	13.8	69.0	1003	US-08-888-077A-14	Sequence 14, Appl
C 27	13.8	69.0	4080	US-08-570-311-7	Sequence 7, Appl

28	13.8	69.0	4080	3	US-08-353-485-7	Sequence 7, Appl
29	13.8	69.0	6241	2	US-08-570-311-25	Sequence 25, Appl
30	13.8	69.0	6895	2	US-08-570-311-9	Sequence 9, Appl
31	13.8	69.0	6895	3	US-08-353-485-9	Sequence 9, Appl
C 32	13.8	69.0	7266	5	US-08-336-308A-9	Sequence 9, Appl
C 33	13.8	69.0	8640	2	US-08-570-311-28	Sequence 28, Appl
C 34	13.8	69.0	12537	3	US-08-611-280-4	Sequence 4, Appl
C 35	13.6	68.0	42	4	US-08-343-443B-49	Sequence 49, Appl
C 36	13.6	68.0	46	4	US-08-381-637-2	Sequence 2, Appl
C 37	13.6	68.0	507	5	US-08-384-106A-3	Sequence 3, Appl
C 38	13.6	68.0	507	6	PCT-US96-01643-3	Sequence 3, Appl
C 39	13.6	68.0	694	4	US-08-627-610-9	Sequence 9, Appl
C 40	13.6	68.0	694	5	US-08-384-106A-4	Sequence 4, Appl
C 41	13.6	68.0	694	6	PCT-US96-01643-4	Sequence 4, Appl
C 42	13.6	68.0	1045	2	US-08-599-252-92	Sequence 92, Appl
C 43	13.6	68.0	1045	6	PCT-US96-06352-92	Sequence 92, Appl
C 44	13.6	68.0	1045	6	PCT-US96-06583-92	Sequence 92, Appl
C 45	13.6	68.0	1674	2	US-08-096-762-199	Sequence 199, App

ALIGNMENTS

RESULT 1  
US-08-211-312-1/c  
; Sequence 1, Application US/08211312  
; Patent No. 5986051  
; GENERAL INFORMATION:  
; APPLICANT: LABIGNE, AGNES  
; APPLICANT: CUSSAC, VALERIE  
; APPLICANT: FERRERO, RICHARD  
; TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY  
; TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/211,312  
; FILING DATE: 01-JUL-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR92/00921  
; FILING DATE: 02-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, NO. 5986051man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 660-075-0XPCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3559 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 2..16
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 211..795
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 800..1309
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1324..2091
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2122..2718
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2721..3515
; US-08-211-312-1

Query Match          77.0%; Score 15.4; DB 4; Length 3559;
Best Local Similarity 94.1%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gaacttcctaaaggga 18
   ||| |||||
Db 717 GAATTCCTAAGGGA 701

RESULT 2
US-08-472-285-1/c
; Sequence 1, Application US/08472285
; Patent No. 6027878
; GENERAL INFORMATION:
; APPLICANT: LABIGNE, AGNES
; APPLICANT: CUSSAC, VALERIE
; APPLICANT: FERRERO, RICHARD
; TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY
; TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,285
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,312
; FILING DATE: 01-JUL-1994
; APPLICATION NUMBER: FR 91 12198
; FILING DATE: 03-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR92/00921
; FILING DATE: 02-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 6027878man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-075-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3559 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..16
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 211..795
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 800..1309
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1324..2091
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2122..2718
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2721..3515
; US-08-472-285-1

Query Match          77.0%; Score 15.4; DB 5; Length 3559;
Best Local Similarity 94.1%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gaacttcctaaaggga 18
   ||| |||||
Db 717 GAATTCCTAAGGGA 701

RESULT 3
US-08-147-777-3/c
; Sequence 3, Application US/08147777
; Patent No. 5914265
; GENERAL INFORMATION:
; APPLICANT: ROOP, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,777
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: two
; APPLICATION NUMBER: 07/876,289
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: Unassigned (204/144)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:

```

```

Query Match      74.0%; Score 14.8; DB 6; Length 24979;
Best Local Similarity 88.9%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ggaacttcctaaaggga 18
      | ||||| |||||
Db      5968 GCAACTTCCCAAGGGA 5951

RESULT      5
US-07-984-044A-1
; Sequence 1, Application US/07984044A
; Patent No. 5461145
; GENERAL INFORMATION:
; APPLICANT: Kudo, T. et al.
; TITLE OF INVENTION: Sexing Method Of Bovine Embryos
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/984,044A
; FILING DATE: 02-DEC-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7005-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-984-044A-1

Query Match      72.0%; Score 14.4; DB 1; Length 1404;
Best Local Similarity 93.8%; Pred. No. 40;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 aacttcctaaaggga 18
      | ||||| |||||
Db      381 AACTTCCCAAGGGA 396

RESULT      6
US-08-458-393-1
; Sequence 1, Application US/08458393
; Patent No. 5661011
; GENERAL INFORMATION:
; APPLICANT: Kudo, T. et al.
; TITLE OF INVENTION: Sexing Method Of Bovine Embryos
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York

```

; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/458,393  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/984,044  
; FILING DATE: 02-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7005-053  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1404 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-458-393-1

Query Match 72.0%; Score 14.4; DB 1; Length 1404;  
Best Local Similarity 93.8%; Pred. No. 40;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 aacttcctcctaaaggga 18  
||||| |||||

Db 381 ANCTTCCCAAAAGGGA 396

RESULT 7  
US-07-865-662F-7  
; Sequence 7, Application US/07865662F  
; Patent No. 5451670  
; GENERAL INFORMATION:  
; APPLICANT: Marcia M. Miller  
; TITLE OF INVENTION: Restriction Fragment Length  
; POLYMORPHISM TEST FOR HAPLOTYPE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: City of Hope  
; STREET: 1500 East Duarte Road  
; CITY: Duarte  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 91010-0269  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3M Double Density 5 1/4" diskette  
; COMPUTER: Wang PC  
; OPERATING SYSTEM: MS DOS Version 3.20  
; SOFTWARE: Microsoft  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/865,662F  
; FILING DATE: 07 April, 1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/688,326  
; FILING DATE: 22 April 1991  
; APPLICATION NUMBER: 07/588,922  
; FILING DATE: 27 September 1990  
; APPLICATION NUMBER: 07/210,405  
; FILING DATE: 23 June 1988

; APPLICATION NUMBER: US 07/130,529  
; FILING DATE: 9 December 1987  
; APPLICATION NUMBER: US 07/068,176  
; FILING DATE: 30 June 1987  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Irons, Edward S.  
; REGISTRATION NUMBER: 16,541  
; REFERENCE/DOCKET NUMBER: No. 5451670e  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 783-6040  
; TELEFAX: (202) 783-6031  
; TELEX: No. 5451670e  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3134  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Double  
; TOPOLOGY: Linear  
; MOLECULE TYPE: DNA  
; ORIGINAL SOURCE: Synthetically Prepared  
; IMMEDIATE SOURCE: Synthetically Prepared  
; US-07-865-662F-7

Query Match 72.0%; Score 14.4; DB 1; Length 3134;  
Best Local Similarity 93.8%; Pred. No. 45;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ctccctcctaaaggagg 20  
||||| |||||

Db 1110 CTCCCTAAAGCGAGG 1125

RESULT 8  
US-08-441-430-1/c  
; Sequence 1, Application US/08441430  
; Patent No. 5681942  
; GENERAL INFORMATION:  
; APPLICANT: Buchwald, Manuel  
; APPLICANT: Strathdee, Craig A.  
; APPLICANT: Wevrick, Rachel  
; APPLICANT: Mathew, Christopher George Porter  
; TITLE OF INVENTION: Fanconi Anemia Type C Gene  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Richard J. Polley, Esq.  
; ADDRESSEE: Klarquist, Sparkman, Campbell, Leigh &  
; ADDRESSEE: Whinston, LLP  
; STREET: 121 S.W. Salmon, Suite 1600  
; CITY: Portland  
; STATE: Oregon  
; COUNTRY: U.S.A.  
; ZIP: 97204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Disk, 3+-inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: WordPerfect 5.1/ASCII Text File  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/441,430  
; FILING DATE: May 15, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. 07/876,285  
; FILING DATE: April 29, 1992  
; APPLICATION NUMBER: U.S. 07/918,313  
; FILING DATE: July 21, 1992  
; APPLICATION NUMBER: U.S. 08/003,963  
; FILING DATE: January 15, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Richard J. Polley, Esq.  
; REGISTRATION NUMBER: 28,107  
; REFERENCE/DOCKET NUMBER: 3812-42824



```

RESULT 12
US-08-901-547A-1
; Sequence 1, Application US/08901547A
; Patent No. 5935837
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Micheal Dolberg
; TITLE OF INVENTION: DNA Constructs And Meth
; TITLE OF INVENTION: Xylose Isomerase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5935837o No. 5935837disk o
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/901,547A
; FILING DATE: 28-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4601.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1668 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,354  
FILING DATE: Concurrent Herewith  
CLASSIFICATION: 0506  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/656,005  
FILING DATE: 24 MAY 1996  
APPLICATION NUMBER: JP 134096/95  
FILING DATE: 31 MAY 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2418/9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5339 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
ORIGINAL SOURCE: Hyperthermophilic archaeon  
US-09-073-354-6

Query Match 71.0%; Score 14.2; DB 5; Length 5339;  
Best Local Similarity 84.2%; Pred. No. 63;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gaacttcctaaaggagg 20  
||||| |||||  
Db 3191 GAACCTCTTTAAAGGATG 3209

RESULT 14  
US-09-073-354-c/c  
; Sequence 6, Application US/09073354  
; Patent No. 6033859  
; GENERAL INFORMATION:  
; APPLICANT: KITABAYASHI, Masao  
; APPLICANT: ARAKAWA, Taku  
; APPLICANT: INOUE, Hiroaki  
; APPLICANT: KAWAKAMI, Bunsei  
; APPLICANT: KAWAMURA, Yoshihisa  
; APPLICANT: IMANAKA, Tadayuki  
; APPLICANT: TAKAGI, Masahiro  
; APPLICANT: MORIKAWA, Masaaki  
; TITLE OF INVENTION: A Thermostable DNA Polymerase and Kits for  
; TITLE OF INVENTION: Amplifying Nucleic Acids  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,354  
FILING DATE: Concurrent Herewith  
CLASSIFICATION: 0506  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/656,005  
FILING DATE: 24 MAY 1996  
APPLICATION NUMBER: JP 134096/95  
FILING DATE: 31 MAY 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2418/9  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5339 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE: Hyperthermophilic archaeon  
US-09-073-354-6

Query Match 71.0%; Score 14.2; DB 5; Length 5339;  
Best Local Similarity 84.2%; Pred. No. 63;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggaacttcctaaaggagg 19  
||| |||||  
Db 5234 GGTCTTCCCTAAAGGAG 5216

RESULT 15  
US-08-902-632-1  
; Sequence 1, Application US/08902632  
; Patent No. 6008025  
; GENERAL INFORMATION:  
; APPLICANT: KOMATSUBARA, Hideyuki  
; APPLICANT: KITABAYASHI, Masao  
; APPLICANT: KAMIMURA, Hideki  
; APPLICANT: KAWAKAMI, Bunsei  
; APPLICANT: KAWAMURA, Yoshihisa  
; APPLICANT: TAKAGI, Masahiro  
; APPLICANT: IMANAKA, Tadayuki  
; TITLE OF INVENTION: Modified Thermostable DNA Polymerase,  
; TITLE OF INVENTION: and DNA Polymerase Composition for Nucleic Acid  
; TITLE OF INVENTION: Amplification  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1 Broadway  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10004

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.50 inch floppy disk, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/902,632  
FILING DATE: Concurrent Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 198911/96  
FILING DATE: 29-JUL-1996  
APPLICATION NUMBER: JP 200446/96  
APPLICATION NUMBER: 30-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: GREASON, Edward W.  
REGISTRATION NUMBER: 18,918  
REFERENCE/DOCKET NUMBER: 2418/7  
TELEPHONE: 212-425-7200  
TELEFAX: 212-425-5288  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5342 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE: hyperthermophilic archaeon  
STRAIN: KOD1  
US-08-902-632-1

Query Match 71.0%; Score 14.2; DB 5; Length 5342;  
Best Local Similarity 84.2%; Pred. No. 63;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gaacttcctaaaggagg 20  
||||| |||||  
Db 3194 GAACCTCTTTAAAGGATG 3212

Search completed: May 23, 2000, 09:40:33  
Job time: 8772 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:40:53 ; Search time 693.2 Seconds  
(without alignments)  
-28.067 Million cell updates/sec

Title: US-08-945-805-4  
Perfect score: 20  
Sequence: 1 aacggcatggactgaatcgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_bal:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pl1:\*
- 8: gb\_pl2:\*
- 9: gb\_pr1:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: gb\_ro:\*
- 13: gb\_sts:\*
- 14: gb\_sy:\*
- 15: gb\_un:\*
- 16: gb\_vi:\*
- 17: em\_fun:\*
- 18: em\_hum1:\*
- 19: em\_hum2:\*
- 20: em\_in:\*
- 21: em\_om:\*
- 22: em\_or:\*
- 23: em\_ov:\*
- 24: em\_pat:\*
- 25: em\_ph:\*
- 26: em\_pl:\*
- 27: em\_ro:\*
- 28: em\_sts:\*
- 29: em\_sy:\*
- 30: em\_un:\*
- 31: em\_vi:\*
- 32: gb\_htg1:\*
- 33: gb\_htg2:\*
- 34: gb\_in1:\*
- 35: gb\_in2:\*
- 36: em\_ba1:\*
- 37: em\_ba2:\*
- 38: em\_hum3:\*
- 39: em\_hum4:\*
- 40: gb\_pr4:\*
- 41: gb\_htg3:\*
- 42: gb\_htg4:\*
- 43: gb\_htg5:\*
- 44: gb\_htg6:\*

- 45: gb\_htg7:\*
- 46: em\_htg1:\*
- 47: em\_htg2:\*
- 48: em\_htg3:\*
- 49: em\_hum5:\*
- 50: gb\_pl3:\*
- 51: gb\_pr5:\*
- 52: gb\_htg8:\*
- 53: gb\_htg9:\*
- 54: gb\_htg10:\*
- 55: gb\_htg11:\*
- 56: gb\_htg12:\*
- 57: gb\_htg13:\*
- 58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	16.4	82.0	171545	51 AC008394	AC008394 Homo sapi
C 2	15.8	79.0	1840	34 D88911	D88911 Fulvia muti
C 3	15.8	79.0	1852	34 D88910	D88910 Vasticardiu
C 4	15.8	79.0	39373	41 AC006625	AC006625 Caenorhab
5	15.8	79.0	123004	42 AC014962	AC014962 Drosophil
6	15.8	79.0	163778	41 AC010868	AC010868 Homo sapi
C 7	15.4	77.0	163	13 G23158	G23158 human STS W
8	15.4	77.0	229	11 HUMYU40B04	AF075109 Homo sapi
9	15.4	77.0	449	1 BACSP00FG	M11081 B.subtilis
10	15.4	77.0	709	13 G55770	G55770 SHGC-100980
11	15.4	77.0	761	1 BSSPOOF	X03497 Bacillus su
C 12	15.4	77.0	1162	1 BSSPOO	V00105 Bacillus su
13	15.4	77.0	2732	5 A58331	A58331 Sequence 3
14	15.4	77.0	2733	40 AF019968	AF019968 Homo sapi
15	15.4	77.0	2805	9 HUMMG44A	L08238 Human MG44
16	15.4	77.0	2820	12 AF019969	AF019969 Mus muscu
17	15.4	77.0	6313	1 BACSP00FA	M22039 Bacillus su
18	15.4	77.0	11570	16 AD5001	X02996 Adenovirus
19	15.4	77.0	25940	1 BSDNA320D	Z49782 B.subtilis
20	15.4	77.0	35935	16 ADRCOMPGEN	M73260 Mastadenovi
21	15.4	77.0	48625	52 AC022214	AC022214 Homo sapi
C 22	15.4	77.0	49625	52 AC022214	AC022214 Homo sapi
C 23	15.4	77.0	70358	41 AC011270	AC011270 Homo sapi
24	15.4	77.0	92854	11 HSJ485C17	AL049845 Human DNA
C 25	15.4	77.0	107102	43 AC008946	AC008946 Homo sapi
C 26	15.4	77.0	107667	32 AF000639	AF000639 Homo sapi
C 27	15.4	77.0	110000	32 CEY116A8_2	Continuation (3 of
C 28	15.4	77.0	110450	41 AC008718	AC008718 Homo sapi
29	15.4	77.0	112595	40 AF196970	AF196970 Homo sapi
C 30	15.4	77.0	115246	41 AC008088	AC008088 Homo sapi
C 31	15.4	77.0	142992	40 AC004838	AC004838 Homo sapi
C 32	15.4	77.0	165423	45 AC019020	AC019020 Homo sapi
C 33	15.4	77.0	184204	44 AC011359	AC011359 Homo sapi
C 34	15.4	77.0	188503	55 AC022596	AC022596 Homo sapi
C 35	15.4	77.0	192973	32 AL133327	AL133327 Homo sapi
C 36	15.4	77.0	200000	43 AC006294	AC006294 Homo sapi
C 37	15.4	77.0	211385	33 AL138881	AL138881 Homo sapi
C 38	15.4	77.0	212150	1 BSUB0020	Z99123 Bacillus su
C 39	15.4	77.0	212610	1 BSUB00019	Z99122 Bacillus su
C 40	15.4	77.0	227950	43 AC015860	AC015860 Homo sapi
C 41	15.4	77.0	235336	43 AC009087	AC009087 Homo sapi
C 42	15.4	77.0	260341	34 CEY116A8C	AL117204 Caenorhab
C 43	15.2	76.0	278	3 RABIGHAP	M21270 Rabbit Ig H
44	15.2	76.0	323	3 AF014688	AF014688 Oryctolag
45	15.2	76.0	349	3 AF058574	AF058574 Oryctolag

ALIGNMENTS

```

RESULT 1
AC008394/c
LOCUS AC008394 171545 bp DNA PRI 02-FEB-2000
DEFINITION Homo sapiens chromosome 5 clone CIT-HSPC_242N15, complete sequence.
ACCESSION AC008394
VERSION AC008394.3 GI:6850300
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 171545)
TITLE Direct Submission
JOURNAL Biol. Bull. 195 (1), 70-77 (1998)
MEDLINE 98411867
FEATURES
source
1..171545
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>171545
rRNA
BASE COUNT 53156 a 32285 c 32391 g 53713 t
ORIGIN
Query Match 82.0%; Score 16.4; DB 51; Length 171545;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 aacggcatggaactgaatc 18
|| |||||
Db 73183 AAGGCATGGACTGAATC 73166

RESULT 2
D88911/c
LOCUS D88911 1840 bp DNA INV 01-SEP-1998
DEFINITION Fulvia mutica 18S rRNA gene, partial sequence.
ACCESSION D88911
VERSION D88911.1 GI:3461878
KEYWORDS 18S ribosomal RNA.
SOURCE Fulvia mutica
ORGANISM Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea;
Cardiidae; Fulvia.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 1840)
TITLE Direct Submission
JOURNAL Submitted (10-NOV-1996) to the DDBJ/EMBL/GenBank databases. Tadashi
Maruyama, Marine Biotechnology Institute, Kamaishi Laboratories;
Heita 3-75-1, Kamaishi, Iwate 026-0001, Japan
(E-mail:tmuruyama@kamaishi.mbio.co.jp, Tel:81-193-26-5814,
Fax:81-193-26-6584)
2 (sites)
Maruyama,T., Ishikura,M., Yamazaki,S. and Kanai,S.
Molecular phylogeny of zooxanthellate bivalves
Biol. Bull. 195 (1), 70-77 (1998)
MEDLINE 98411867
FEATURES
source
1..1840
/organism="Fulvia mutica"
/db_xref="taxon:80828"
<1..>1840
rRNA
BASE COUNT 433 a 447 c 535 g 425 t
ORIGIN
Query Match 79.0%; Score 15.8; DB 34; Length 1840;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacggcatggaactgaatcg 19
|| |||| |
Db 1688 AAGGCAGGACTGAATCG 1670

RESULT 3
D88910/c
LOCUS D88910 1852 bp DNA INV 01-SEP-1998
DEFINITION Vastocardium flavum 18S rRNA gene, partial sequence.
ACCESSION D88910
VERSION D88910.1 GI:3461877
KEYWORDS 18S ribosomal RNA.
SOURCE Vastocardium flavum
ORGANISM Vastocardium flavum
Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea;
Cardiidae; Vastocardium.
REFERENCE
AUTHORS 1 (bases 1 to 1852)
TITLE Direct Submission
JOURNAL Submitted (10-NOV-1996) to the DDBJ/EMBL/GenBank databases. Tadashi
Maruyama, Marine Biotechnology Institute, Kamaishi Laboratories;
Heita 3-75-1, Kamaishi, Iwate 026-0001, Japan
(E-mail:tmuruyama@kamaishi.mbio.co.jp, Tel:81-193-26-5814,
Fax:81-193-26-6584)
2 (sites)
Maruyama,T., Ishikura,M., Yamazaki,S. and Kanai,S.
Molecular phylogeny of zooxanthellate bivalves
Biol. Bull. 195 (1), 70-77 (1998)
MEDLINE 98411867
FEATURES
source
1..1852
/organism="Vastocardium flavum"
/db_xref="taxon:80826"
<1..>1852
rRNA
BASE COUNT 429 a 444 c 548 g 431 t
ORIGIN
Query Match 79.0%; Score 15.8; DB 34; Length 1852;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacggcatggaactgaatcg 19
|| |||| |
Db 1701 AAGGCAGGACTGAATCG 1683

```

```

RESULT 4
AC006625/c
LOCUS
DEFINITION Caenorhabditis elegans clone C55B7, *** SEQUENCING IN PROGRESS ***
ACCESSION AC006625
VERSION AC006625.1 GI:4263313
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
AUTHORS Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE 1 (bases 1 to 39373)
JOURNAL Waterston,R.H.
AUTHORS The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39373)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 39373: contig of 39373 bp in length.
*
FEATURES
source
Location/Qualifiers
1..39373
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="C55B7"
BASE COUNT 12580 a 6712 c 7027 g 13054 t
ORIGIN
Query Match 79.0%; Score 15.8; DB 41; Length 39373;
Best Local Similarity 89.5%; Pred. No. 5.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 acggcatgactgaatcgg 20
|||||
Db 29603 ACGGCATGGACATATCGG 29585
|||||

RESULT 5
AC014962
LOCUS
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION AC014962
VERSION AC014962.1 GI:6436373
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 123004)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10211724 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced

```

```

* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
Location/Qualifiers
1..123004
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 36645 a 24655 c 24662 g 37042 t
ORIGIN
Query Match 79.0%; Score 15.8; DB 42; Length 123004;
Best Local Similarity 89.5%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 acggcatgactgaatcgg 20
|||||
Db 89024 ACGGCATGGACGAATCGG 89042
|||||

RESULT 6
AC010868
LOCUS
DEFINITION Homo sapiens chromosome 15 clone BAC 396D15 map 15q24, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC010868
VERSION AC010868.1 GI:5923650
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163778)
AUTHORS Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
TITLE Sequencing of human chromosome 15 D15S160-D15S115 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163778)
AUTHORS Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT
* NOTE: This record contains 192 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 857: contig of 857 bp in length
* gap of unknown length
* 858 1747: contig of 890 bp in length
* gap of unknown length
* 1748 2588: contig of 841 bp in length
* gap of unknown length
* 2589 3423: contig of 835 bp in length
* gap of unknown length
* 3424 4247: contig of 824 bp in length
* gap of unknown length
* 4248 5097: contig of 850 bp in length
* gap of unknown length
* 5098 5994: contig of 897 bp in length
* gap of unknown length
* 5995 6865: contig of 871 bp in length
* gap of unknown length
* 6866 7830: contig of 965 bp in length
* gap of unknown length

```

* 7831	8710:	contig of 880 bp in length	* 39884	40703:	gap of unknown length
* *	9538:	gap of unknown length	* *	41582:	contig of 820 bp in length
* *	10377:	contig of 828 bp in length	* *	42408:	gap of unknown length
* *	11198:	contig of 839 bp in length	* *	43245:	gap of unknown length
* *	12035:	contig of 821 bp in length	* *	44069:	gap of unknown length
* *	12882:	contig of 837 bp in length	* *	44915:	contig of 824 bp in length
* *	13728:	contig of 847 bp in length	* *	45824:	contig of 846 bp in length
* *	14554:	contig of unknown length	* *	47489:	gap of unknown length
* *	15388:	contig of 825 bp in length	* *	48339:	contig of 909 bp in length
* *	16218:	contig of 834 bp in length	* *	49160:	contig of 842 bp in length
* *	17062:	contig of 830 bp in length	* *	50004:	gap of unknown length
* *	18358:	contig of 844 bp in length	* *	50847:	contig of 844 bp in length
* *	20669:	contig of unknown length	* *	51686:	gap of unknown length
* *	21515:	contig of 819 bp in length	* *	52527:	contig of 839 bp in length
* *	22340:	contig of 846 bp in length	* *	53372:	gap of unknown length
* *	23184:	contig of 825 bp in length	* *	54239:	contig of 845 bp in length
* *	24010:	contig of 844 bp in length	* *	55088:	gap of unknown length
* *	24852:	contig of 826 bp in length	* *	55910:	contig of 867 bp in length
* *	25683:	contig of unknown length	* *	56779:	gap of unknown length
* *	26576:	contig of 893 bp in length	* *	57608:	contig of 822 bp in length
* *	27399:	contig of 842 bp in length	* *	58461:	gap of unknown length
* *	28245:	contig of 823 bp in length	* *	59284:	gap of unknown length
* *	29068:	contig of 846 bp in length	* *	60121:	contig of 823 bp in length
* *	29909:	contig of unknown length	* *	61017:	contig of 837 bp in length
* *	30730:	contig of 841 bp in length	* *	61893:	contig of 896 bp in length
* *	31572:	contig of 821 bp in length	* *	62721:	gap of unknown length
* *	32392:	contig of 842 bp in length	* *	63563:	contig of 828 bp in length
* *	33228:	contig of 820 bp in length	* *	64435:	gap of unknown length
* *	34065:	contig of 836 bp in length	* *	65275:	contig of 872 bp in length
* *	34900:	contig of 837 bp in length	* *	66095:	gap of unknown length
* *	35722:	contig of 835 bp in length	* *	66932:	contig of 840 bp in length
* *	36591:	contig of unknown length	* *	67768:	gap of unknown length
* *	37413:	contig of 822 bp in length	* *	68617:	contig of 836 bp in length
* *	38253:	contig of 849 bp in length	* *	69462:	gap of unknown length
* *	39081:	contig of 840 bp in length	* *	70304:	contig of 845 bp in length
* *	39883:	contig of 828 bp in length	* *		contig of 842 bp in length
* *		gap of unknown length	* *		gap of unknown length

\* 70305 71131: contig of 827 bp in length  
gap of unknown length

Query Match 79.0%; Score 15.8; DB 41; Length 163778;  
Best Local Similarity 89.5%; Pred. No. 6e+02; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 2;

Oy 2 acggcatggactgaatcgg 20  
|||||  
Db 97049 ACAGCTGGACTGAATCGG 97067

RESULT 7  
LOCUS G23158 163 bp DNA STS 31-MAY-1996  
DEFINITION human STS WI-18178, sequence tagged site.  
ACCESSION G23158  
VERSION G23158.1 GI:1343484  
KEYWORDS STS; STS sequence; primer; sequence tagged site.  
SOURCE human STSs derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 163)  
AUTHORS Hudson,T.  
TITLE Whitehead Institute/MIT Center for Genome Research; Physically  
Mapped STSs  
JOURNAL Unpublished (1995)  
COMMENT Contact: Thomas Hudson  
Whitehead Institute/MIT Center for Genome Research  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA  
Tel: 617 252 1900  
Fax: 617 252 1902  
Email: thudson@genome.wi.mit.edu

Primer A: GCATAGGTTGAGGGGTGA  
Primer B: CCGTATGCCCTGCTATCAGATG  
STS size: 128  
PCR Profile:  
Presoak:  
Denaturation:  
Annealing: 56 degrees C  
Polymerization:  
PCR Cycles: 35  
Thermal Cycler:  
Protocol:  
Template: 10 ng  
Primer: each 5 pM  
dNTPs: each 4 mM  
Taq Polymerase: 0.025 units/ul  
Total Vol: 20 ul

Buffer:  
MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCL: 10 mM  
pH: 9.3

Derived from dbEST (genbank accession H61188).

FEATURES  
source  
1. .163  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="91.8 CR from top of ChrX linkage group"  
17. .144  
17. .36  
STS primer\_bind complement(123..144)  
primer\_bind 41 a 36 c 54 g 32 t  
BASE COUNT  
ORIGIN

Query Match 77.0%; Score 15.4; DB 13; Length 163;  
Best Local Similarity 94.1%; Pred. No. 8.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 ggcacggactgaatcgg 20  
|||||  
Db 65 GGCATGGACTGAATCTG 49

RESULT 8  
LOCUS HUMYU40D04 229 bp mRNA PRI 04-AUG-1998  
DEFINITION Homo sapiens full length insert cDNA YU40D04.  
ACCESSION AF075109  
VERSION AF075109.1 GI:3377650  
KEYWORDS FLI\_CDNA.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 229)  
AUTHORS Woessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,  
Marth,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,  
Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,  
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,  
Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behymer,K., Hillier,L.,  
Wilson,R. and Waterston,R.

TITLE Full Clone Sequencing of the Longest Available Member from Each  
Unigene Cluster  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 229)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUN-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT SUBMITTED BY:  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

FEATURES  
source  
1. .229  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares ovary\_tumor\_NbHOT"  
/clone="IMAGE:236263"  
56 a 70 c 46 g 57 t

BASE COUNT  
ORIGIN

Query Match 77.0%; Score 15.4; DB 11; Length 229;  
Best Local Similarity 94.1%; Pred. No. 8.3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 ggcacggactgaatcgg 20  
|||||

Db 133 GGCATGGACTGAATCG 149

# RESULT 9

LOCUS BACSP00FG 449 bp DNA BCT 26-APR-1993  
 DEFINITION B.subtilis spo0F 2 gene coding for early sporulation protein spo0F.  
 ACCESSION M11081  
 VERSION M11081.1 GI:143600  
 KEYWORDS early sporulation protein; spo0F gene.  
 SOURCE B.subtilis (168T) DNA, clones pJH4122 and pJH4133.  
 ORGANISM Bacillus subtilis  
 Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/Staphylococcus group; Bacillus.  
 REFERENCE 1 (bases 1 to 449)  
 AUTHORS Trach,K.A., Chapman,J.W., Piggot,P.J. and Hoch,J.A.  
 TITLE Deduced product of the stage 0 sporulation gene spo0F shares  
 homology with the spo0A, OmpR, and SfrA proteins  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 82, 7260-7264 (1985)  
 MEDLINE 86042645  
 COMMENT Draft entry and computer-readable sequence in [1] kindly provided  
 by J.A.Hoch, 23-JAN-1986.  
 The spo0F locus defined in [1] is different from a previously  
 sequenced putative spo0F locus (Shimotsu, et al., 1983; see  
 separate entry).  
 A putative ribosome binding site is present at positions 22-27.

# FEATURES

source  
 1..449  
 /organism="Bacillus subtilis"  
 /sub\_strain="PY79"  
 /strain="W168"  
 /db\_xref="taxon:1423"  
 32..406  
 /note="Spo0F protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AAA22787.1"  
 /db\_xref="GI:143601"  
 /translation="MNNKILIVDDYGRILLNEVFNKEGYQTQAAAGLQALDIVT  
 KERPLVLLDKIPGMDGIEIKRMKVIDENIRVIIMTAIGELDIQSKELGALTHF  
 AAFPDIIDEIRDAVKRYLPLKSN"  
 51  
 mutation  
 /note="t in wild type; a in mutant spo0F221"  
 BASE COUNT 149 a 87 c 106 g 107 t  
 ORIGIN 61 bp upstream of BclI site.

Query Match 77.0%; Score 15.4; DB 1; Length 449;  
 Best Local Similarity 94.1%; Pred. No. 8.4e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19

|||||||

Db 205 CGCATGGACGAATCG 221

# RESULT 10

LOCUS G55770 709 bp DNA STS 26-OCT-1999  
 DEFINITION SHGC-100980 Human Homo sapiens STS genomic, sequence tagged site.  
 ACCESSION G55770  
 VERSION G55770.1 GI:6120939  
 KEYWORDS STS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 709)  
 AUTHORS Myers,R.M.  
 TITLE Human STSs (1999)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Richard M. Myers

Stanford Human Genome Center (SHGC)  
 Stanford University School of Medicine  
 Department of Genetics, M-344, Stanford, CA 94305, USA  
 Tel: 4157259687  
 Fax: 4157259689

Email: myers@shgc.stanford.edu

Primer A: TAGTTCACACATTGCCACAGAA

Primer B: AGATGGTGGGAAGTAACAGGT

STS size: 325

PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds

Annealing: 60 degrees C for 30 seconds

Polymerization: 72 degrees C for 23 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

AmpliTaq Gold Polymerase: 0.07 units/ul

Total Vol: 5 ul

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed  
 and developed at the Stanford Human Genome Center.

# FEATURES

source  
 1..709  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="7"  
 /clone\_lib="Human"  
 72..396  
 primer\_bind 72..94  
 primer\_bind complement(374..396)  
 BASE COUNT 254 a 131 c 107 g 217 t  
 ORIGIN

Query Match 77.0%; Score 15.4; DB 13; Length 709;  
 Best Local Similarity 94.1%; Pred. No. 8.5e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggcattgactgaatcgg 20

|||||||

Db 163 GGCATGGACTGAATCTG 179

# RESULT 11

LOCUS BSSPOOF 761 bp DNA BCT 12-SEP-1993  
 DEFINITION Bacillus subtilis spo0F gene.  
 ACCESSION X03497  
 VERSION X03497.1 GI:40155  
 KEYWORDS regulatory protein; sporulation; unidentified reading frame.  
 SOURCE Bacillus subtilis.  
 ORGANISM Bacillus subtilis  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;  
 Bacillus.  
 REFERENCE 1 (bases 1 to 761)  
 AUTHORS Yoshikawa,H., Kazami,J., Yamashita,S., Chibazakura,T., Sone,H.,  
 Kawamura,F., Oda,M., Isaka,M., Kobayashi,Y. and Saito,H.  
 TITLE Revised assignment for the Bacillus subtilis spo0F gene and its  
 homology with spo0A and with two Escherichia coli genes  
 JOURNAL Nucleic Acids Res. 14 (2), 1063-1072 (1986)  
 MEDLINE 86120355  
 FEATURES  
 Location/Qualifiers  
 1..761  
 source

```
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
complement(<1..81)
/note="0.5 kb gene (aa 1-27)"
/codon_start=1
/transl_table=11
/protein_id="CAA27216.1"
/db_xref="GI:40156"
/db_xref="SWISS-PROT:P06629"
/translation="MLKIFTQLTGIFSRQDKESDALEDG"
204
/note="transcription start"
227..236
/note="imp. Inverted repeat A"
237..242
/note="put. ribosome binding site"
240
/note="g is t in spoOF 77"
244..253
/note="imp. Inverted repeat A'"
247..621
/note="spoOF protein (aa 1-124)"
/codon_start=1
/transl_table=11
/protein_id="CAA27217.1"
/db_xref="GI:40157"
/db_xref="SWISS-PROT:P06628"
/translation="MMNEKILIVDDOYGIRILLNEVFNKGYQTFQANGLOALDIYV
KERPDVLDMKIPGMDGIEILKRMKVIDENIRVIITAYGELDMIQESKELGALTFF
AKPFDIIDEIRDAVKKIPLKSN"
263
mutation
265 a 138 c 162 g 196 t
/note="t is a (ile is Asp) in spoOF 221"
BASE COUNT 265 a 138 c 162 g 196 t
ORIGIN

Query Match 77.0%; Score 15.4; DB 1; Length 761;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19
|||||
Db 420 CGGCATGGACGGATCG 436

RESULT 12
BSSPOO/c 1162 bp DNA BCT 12-SEP-1993
LOCUS Bacillus subtilis gene required at an early stage of sporulation.
DEFINITION (gene code spoOF).
ACCESSION V00105 J01549
VERSION V00105.1 GI:40176
KEYWORDS sporulation.
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
REFERENCE 1 (bases 1 to 1162)
AUTHORS Shimotsu,H., Kawamura,F., Kobayashi,Y. and Saito,H.
TITLE Early sporulation gene spoOF: nucleotide sequence and analysis of
gene product
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80 (3), 658-662 (1983)
MEDLINE 83143990
FEATURES Location/Qualifiers
1..1162
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
541..1062
/gene="spoOF"
541..1062
/gene="spoOF"
/codon_start=1
/transl_table=11
```

```
/protein_id="CAA23438.1"
/db_xref="GI:40177"
/note="SWISS-PROT:P06629"
/translation="MLKIFTQLTGIFSRQDKESDALEDGRLAQAIVISHSTIYLY
GANELOGVFYEATESKEPPFSPVKAFFENAEVTSRDLVMEFCSTGTAEQELAKELY
EKAGVVCVSPAARDSAGIEQICDVHDISKMLPLVDEDEGTRYGFPSLMTALVYHA
LSFTLKEILQEYA"
BASE COUNT 308 a 246 c 253 g 355 t
ORIGIN

Query Match 77.0%; Score 15.4; DB 1; Length 1162;
Best Local Similarity 94.1%; Pred. No. 8.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19
|||||
Db 203 CGGCATGGACGGATCG 187

RESULT 13
A58331
LOCUS A58331 2732 bp DNA PAT 05-MAR-1998
DEFINITION Sequence 3 from Patent WO9635784.
ACCESSION A58331
VERSION A58331.1 GI:3713993
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2732)
Jenuwein,T. and Laible,G.
CHROMATIN-REGULATOR GENES
Patent: WO 9635784-A 3 14-NOV-1996;
BOHRINGER INGELHEIM INT (DE)
Other publication DE 19516776 961114.
FEATURES Location/Qualifiers
1..2732
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="B-CELL"
1..44
45..1283
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA03483.1"
/db_xref="GI:3713994"
/translation="MAENLKGSVCCKSSWNQLODLRLAKLSCPALGISRKNLYDFE
VEYLDYKKIREQYVLYVQKORRALRWEQELNKRSHLGRITVENEVDLDGPPRAF
RSTPRHLDPSLANLYVQKORRALRWEQELNKRSHLGRITVENEVDLDGPPRAF
VYINEYRVGEITLNOVAVGCEQDCLWAPTGCCPGASLHRFAYNDGQVRLRAGLP
IITECSRRCGYDCPNRVQGIYDLCIFRTDDBRGWGVRTLEKIRKNSFVMEYVE
IITSEAEARGQIYDRQATYLFDLVDYEDVYTDAAVYGNISHFVNHSCDPNLQVYN
VFIDNLERLPRIAFFATRTIRAGELTFDYNMQVDPVDMESTRMSDFGLAGLPSP
KKRYIECKCGTESCRKYLFP"
3'UTR 1284..2732 789 g 566 t
BASE COUNT 595 a 782 c 789 g 566 t
ORIGIN

Query Match 77.0%; Score 15.4; DB 5; Length 2732;
Best Local Similarity 94.1%; Pred. No. 8.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggcattgactgaatcg 20
|||||
Db 2636 GGCATGGACTGAATCG 2652

RESULT 14
AF019968
LOCUS AF019968 2733 bp mRNA PRI 16-APR-1999
```



• • • • •

Tue May 23 16:14:16 2000

us-08-945-805-4.rge

Page 9

-----

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:42:13 ; Search time 88.1 Seconds  
(without alignments)  
56.797 Million cell updates/sec

Title: US-08-945-805-4

Perfect score: 20

Sequence: 1 aacggcatgactgaatcgg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.4	77.0	128	X12310	Human biallelic po
2	15.4	77.0	147	T26163	Human gene signatu
3	15.4	77.0	2732	T43625	Chromatin regulato
4	15.4	77.0	32026	T60559	Recombinant adenov
5	15.4	77.0	34303	T07261	Adenoviral vector
6	15.4	77.0	34382	X15627	Recombinant adenov
7	15.4	77.0	34427	X07371	Adenovirus vector
8	15.4	77.0	35000	T60557	Recombinant cis-ac
9	15.4	77.0	35408	T59272	Recombinant adenov
10	15.4	77.0	35935	T07258	Adenovirus 5 genom
11	15.4	77.0	36538	T60558	Recombinant trans-
12	14.8	74.0	2220	Q33008	Novel intestinal o
13	14.4	72.0	24593	N50226	Sequence of opine
14	14.4	72.0	24596	N50182	Complete nucleotid
15	14.2	71.0	483	T67032	Salmonella iagb ge
16	14.2	71.0	858	X30354	DNA encoding a hum
17	14.2	71.0	909	T33965	Inosine-guanosine
18	14.2	71.0	1209	V74286	R. eutropha Mgt OR
19	14.2	71.0	1635	Q79080	Frtl gene having m
20	14.2	71.0	1713	T50834	Human chitinase cd
21	14.2	71.0	1894	T77328	Class II EPSs DNA
22	14.2	71.0	1894	T93795	Class II EPS synt
23	14.2	71.0	1894	V58020	Synechocystis sp.
24	14.2	71.0	2344	T89885	Drosophila frizzole
25	14.2	71.0	2498	V40801	Soybean raffinose
26	13.8	69.0	22	T63189	FSH receptor fshr
27	13.8	69.0	453	V84059	CDNA encoding a cy
28	13.8	69.0	553	X51720	DNA encoding a hum
29	13.8	69.0	1756	Q31326	E. necatrix srrRNA
30	13.8	69.0	1756	Q31328	E. tenella srrRNA.
31	13.8	69.0	1756	Q31327	Eimeria necatrix s
32	13.8	69.0	2002	N92273	Cloned gene contg.
33	13.8	69.0	2521	T47252	Bacillus subtilis
34	13.8	69.0	2529	N92274	Sequence of HindII

C 35	13.8	69.0	2861	1	T28639	RNAse L inhibitor
C 36	13.8	69.0	3568	1	T28638	RNAse L inhibitor
C 37	13.8	69.0	36335	1	Q68003	Ad2/-ORF6/PK-CFTR
C 38	13.6	68.0	293	1	T23590	Human gene signatu
C 39	13.6	68.0	294	1	V68619	Nucleotide sequenc
C 40	13.6	68.0	508	1	Q31287	Fragment 4, a PCR
C 41	13.6	68.0	508	1	Q31639	E. brunetti probe
C 42	13.6	68.0	527	1	V87894	EST clone FG380. N
C 43	13.6	68.0	1747	1	Q94231	Bovine Neospora is
C 44	13.6	68.0	1747	1	V00075	Neospora nuclear s
C 45	13.6	68.0	2273	1	Q04339	Aquaricine I gene

ALIGNMENTS

RESULT 1

X12310/c	1
ID X12310 standard; DNA; 128 BP.	
AC X12310;	
DT 30-MAR-1999 (first entry)	
DE Human biallelic polymorphic DNA fragment WI-18178.	
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;	
KW detection; phenotypic typing; characteristic; infection; hereditary;	
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;	
KW treatment; marker; ss.	
OS Homo sapiens.	
PN W09820165-A2.	
PD 14-MAY-1998.	
PF 05-NOV-1997; U20313.	
PR 06-NOV-1996; US-030455.	
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.	
PI Hudson T. Lander ES, Wang D;	
DR WPL; 98-286974/25.	
PT New isolated nucleic acid segments from the human genome - used for	
PT determining polymorphic forms for use in e.g. forensics, paternity	
PT testing or phenotypic typing for disease	
PS Claim 1; Page 241; 310pp; English.	
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic	
CC markers which have been isolated using the primers represented in	
CC X09121-X10268. The base occupying the polymorphic site is indicated by	
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in	
CC methods for determining polymorphic forms in an individual for use in	
CC e.g. forensics, paternity testing or for phenotypic typing for diseases	
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,	
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial	
CC hypercholesterolemia, polycystic kidney disease, hereditary	
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, Ehlers-Danlos	
CC haemorrhagic telangiectasia, familial colonic polyposis, acute intermittent porphyria,	
CC syndrome, osteogenesis imperfecta, cancer, diseases of the nervous	
CC autoimmune diseases, inflammation, infection by pathogenic microorganisms, and characteristics such	
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,	
CC endurance, fertility, and susceptibility or receptivity to particular	
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid	
CC segments can also be used to produce medicaments for the treatment or	
CC prophylaxis of such diseases.	
CC Sequence 128 BP; 35 A; 24 C; 46 G; 22 T;	
SQ	

Query Match 77.0%; Score 15.4; DB 1; Length 128;  
Best Local Similarity 94.1%; Pred. No. 16;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggcattgactgaatcgg 20

|||||||

Db 49 GGCATGGACTGAATCTG 33

RESULT 2

T26163

ID T26163 standard; cDNA to mRNA; 147 BP.

AC T26163;

DT 18-OCT-1996 (first entry)  
 DE Human gene signature HUMGS08400.  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 OS Homo sapiens.  
 PN WO9514772-A1.  
 PD 01-JUN-1995.  
 PF 11-NOV-1994; J01916.  
 PR 12-NOV-1993; JP-355504.  
 PA (MATSU) MATSUBARA K.  
 PA (OKUBU) OKUBO K.  
 PI Matsubara K, Okubo K;  
 DR WPI: 95-206931/27.  
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1: Page 2017; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-R26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 147 BP; 35 A; 45 C; 30 G; 36 T;

Query Match 77.0%; Score 15.4; DB 1; Length 147;  
 Best Local Similarity 94.1%; Pred. No. 17;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 ggcattgactgaatcg 20  
 |||||  
 Db 55 GGCATGGACTGAATCTG 71

RESULT 3  
 T43625  
 ID T43625 standard; DNA; 2732 BP.  
 AC T43625;  
 DT 05-MAY-1997 (first entry)  
 DE Chromatin regulator protein SUV39H.  
 KW Chromatin; regulator; EZH1; EZH2; SUV39H; SET domain;  
 KW gene therapy; cancer; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 45..1283  
 FT /\*tag= a  
 PN WO9635784-A2.  
 PD 14-NOV-1996.  
 PF 02-MAY-1996; E01818.  
 PR 10-MAY-1995; DE-016776.  
 PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.  
 PI Jenuwein T, Laible G;  
 DR WPI: 96-518672/51.  
 DR P-PSDB: W05261.  
 PT New DNA encoding chromatin regulator protein with SET domain - and  
 PT related vectors; transformed cells; proteins and antibodies, for  
 PT diagnosis and treatment of cancer  
 PS Claim 1: Fig 6; 38pp; German.  
 CC The DNA was isolated by screening a human B cell cDNA library with  
 CC mixed Drosophila DNA probes based on the conserved SET domains  
 CC in E(z) and Su(var)3-9. The DNA, and its products, are useful

CC in therapy (esp. gene therapy) and diagnosis of human diseases  
 CC that involve deregulated chromatin-regulator genes having a SET  
 CC domain, esp. cancer.  
 SQ Sequence 2732 BP; 595 A; 782 C; 789 G; 566 T;

Query Match 77.0%; Score 15.4; DB 1; Length 2732;  
 Best Local Similarity 94.1%; Pred. No. 26;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 ggcattgactgaatcg 20  
 |||||  
 Db 2636 GGCATGGACTGAATCTG 2652

RESULT 4  
 T60559  
 ID T60559 standard; DNA; 32026 BP.  
 AC T60559;  
 DT 05-MAR-1998 (first entry)  
 DE Recombinant adenovirus H5.020TKneo-Int.  
 KW Adenovirus H5.020TKneo-Int; plasmid pADMLVneo-Int; Ad5; Mo-MLV;  
 KW retrovirus; retrotransposition; transposition; transgene;  
 KW gene therapy; vector; neomycin resistance; neo gene; ss.  
 OS Chimeric - Mastadenovirus serotype 5.  
 OS Chimeric - Synthetic.  
 PN WO9715679-A1.  
 PD 01-MAY-1997.  
 PF 24-OCT-1996; U17176.  
 PR 27-OCT-1995; US-005942.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 PI Kelley WM, Wilson JM;  
 DR WPI: 97-259031/23.  
 PT Recombinant replication defective virus - Inserts transgene into  
 PT host cell chromatin in the presence of transposase, providing stable  
 PT and durable transgene expression  
 PS Example 2: Fig 10A-H; 74pp; English.  
 CC This is the DNA sequence of the recombinant adenovirus  
 CC H5.020TKneo-Int. It was obtained by co-transfecting HEK293 (ATCC  
 CC CRL1573) cells with linearised plasmid pADneo-Int. (see T60555) and  
 CC Ad5dl7001. It was used as a control in studies of the  
 CC retrotransposition and integration of cis- and trans-acting  
 CC recombinant adenoviruses (see T60557 and T60558) into host cells.  
 CC Claimed novel recombinant replication defective viruses (A)  
 CC comprise: (a) DNA of, or corresponding to, at least part of the  
 CC viral genome, able to infect a mammalian cell; and (b) a first  
 CC expression sequence (ES1) comprising a human gene (I) linked to  
 CC the cis-acting terminal repeat sequences of a transposon, and ES1  
 CC flanked by DNA of (a); (A) can infect a mammalian cell and can  
 CC express (I) and transfer it to the cellular chromatin in vivo or in  
 CC vitro in the presence of a transposase. Also new are: (1) a  
 CC recombinant replication defective virus (B) containing (a) as above  
 CC and a second expression sequence (ES2) containing a trans-acting  
 CC transposase gene (II) plus regulatory sequences, flanked by DNA of  
 CC (a) and able to express transposase in mammalian cells; and (2)  
 CC mammalian cells which stably express (I) integrated into its  
 CC chromatin, produced by infection with (A). (A) and (B) are used  
 CC for delivery and stable integration of (I) into a host cell  
 CC (claimed), e.g. in somatic gene therapy of genetic defects or  
 CC deficiencies such as cystic fibrosis. (A) are also used to prepare  
 CC recombinant retroviruses (RRV) (claimed). Because (I) becomes  
 CC stably integrated, it provides longer lasting expression than genes  
 CC introduced with conventional adenoviral vectors and the need for  
 CC repeated administration is avoided. The transgene is inherited by  
 CC progeny cells and the viral component is gradually degraded by the  
 CC cell. When used for RRV production, (A) increase the ratio of RRV  
 CC to empty retrovirus.  
 SQ Sequence 32026 BP; 7388 A; 9141 C; 8873 G; 5624 T;

Query Match 77.0%; Score 15.4; DB 1; Length 32026;  
 Best Local Similarity 94.1%; Pred. No. 39;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19  
 ||||| |||||  
 Db 10521 CGCATGGCTGAATCG 10537

## RESULT 5

ID V07261 standard; DNA; 34303 BP.  
 AC V07261;  
 DT 28-SEP-1998 (first entry)  
 DE Adenoviral vector plasmid pBHGl1.  
 KW Adenovirus 5; Ad5; vector; gene therapy; ds.  
 OS Mastadenovirus 5.  
 FH Key Location/Qualifiers  
 FT misc\_feature 8772..9385  
 FT /\*tag= a  
 FT /note= "region deleted in plasmid pdePTP and  
 pBHGl1delPTP (Claim 45)"  
 FT misc\_feature 8773..9586  
 FT /\*tag= b  
 FT /note= "region deleted in plasmid  
 pAXdelpoldelPTPVARNA+tl3 and  
 pBHGl1delpolydelPTPVARNA+tl3 (Claim 53)"  
 FT misc\_feature 10705..11134  
 FT /\*tag= c  
 FT /note= "region deleted in plasmid pdePTP and  
 pBHGl1delPTP (Claim 51)"  
 FT misc\_feature 11067..12513  
 FT /\*tag= d  
 FT /note= "region deleted in plasmid  
 pAXdelpoldelPTPVARNA+tl3 and  
 pBHGl1delpolydelPTPVARNA+tl3 (Claim 53)"

WO9817783-A1.

PN 30-APR-1998.  
 PD 23-OCT-1997; U19541.  
 PF 23-OCT-1997; US-735609.  
 PR (UNMI ) UNIV MICHIGAN.  
 PA Analifitano A, Chamberlain JS, Hartigan-O'Connor DJ,  
 PI Hauser MA, Kumar-Singhr;  
 PI WPI; 98-261485/23.  
 DR New adenoviral recombinant plasmid(s) - comprise sequences provided  
 PT for expression of large foreign DNA fragments, used for, e.g. gene  
 PT therapy of genetic diseases(s).  
 PS Claim 45; Page 87-102; 139pp; English.  
 CC This nucleotide sequence comprises plasmid pBHGl1 that consists of  
 CC an EI-deleted adenovirus (Ad) genome. It contains a deletion of  
 CC Ad5 (see also V07258) from bp 188-1339 (0.5-3.7 m.u.). This  
 CC deletion removes the packaging signals as well as the E1 sequences.  
 CC pBHGl1 also contains a large deletion within the E1 region (bp  
 CC 27865-30995, 77.5-86.2 m.u.). Nucleotide 8773 of pBHGl1 is  
 CC equivalent to nucleotide 7269 of Ad5. The large E3 deletion  
 CC provides a large cloning capacity to Ad vectors based on pBHGl1.  
 CC Recombinant plasmids pdePTP and pBHGl1delPTP (deleted for E1 and  
 CC polymerase functions), and pAXdelpoldelPTPVARNA+tl3 and  
 CC pBHGl1delpoldelPTPVARNA+tl3PBHGl1 (deleted for E1, polymerase as  
 CC preterminal protein functions), are specifically claimed. The  
 CC invention provides improved adenoviral vectors and packaging  
 CC cell lines. One type of improved vector comprises deletions  
 CC within the E2b region of the adenoviral genome (see also V07261).  
 CC These E2b-deleted virus are used in conjunction with novel cell  
 CC lines that constitutively express E2b gene products. The invention  
 CC also provides vectors deleted for all viral coding regions. These  
 CC 'guttred' vectors permit the transfer of large genes (e.g. up to  
 CC 35 kb) to cells, as demonstrated by the transfer of the dystrophin  
 CC gene to the muscle of mice. The E2b-deleted and gutted vectors  
 CC provide improved adenoviral vectors useful for a wide variety of  
 CC gene therapy applications.

Query Match 77.0%; Score 15.4; DB 1; Length 34303;

Sequence 34303 BP; 7948 A; 9625 C; 9517 G; 7213 T;

Best Local Similarity 94.1%; Pred. No. 39;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19  
 ||||| |||||  
 Db 12914 CGCATGGCTGAATCG 12930

## RESULT 6

ID X15627 standard; DNA; 34382 BP.  
 AC X15627;  
 DT 07-MAY-1999 (first entry)  
 DE Recombinant adenovirus Ad:Pac-beta-Gal.  
 KW Recombinant adenovirus Ad:Pac-beta-Gal; replication deficient adenovirus;  
 KW Ela region; E3 region; gene therapy; nitric oxide synthetase; NOS;  
 KW cystic fibrosis chloride channel; LDL receptor; erythropoietin;  
 KW atherosclerotic artery; ss.  
 OS Synthetic.  
 OS Mastadenovirus.  
 FH Key Location/Qualifiers  
 FT CDS complement (3372..333)  
 FT /\*tag= a  
 FT US5880102-A.  
 PD 09-MAR-1999.  
 PF 17-JAN-1995; 374483.  
 PR 17-JAN-1995; US-374483.  
 PA (OYDU-) UNIV DUKE.  
 PI Blazing MA, George SE;  
 DR WPI; 99-204005/17.  
 DR P-PSDB; W97243.  
 DT New replication deficient adenovirus bearing deletions of the Ela  
 PT and E3 regions - containing a single packaging signal sequence and  
 PT Ela enhancer sequence, the Ela deletion has unique cleavage sites  
 PT and is useful as a gene therapy vector  
 PS Disclosure; Fig 49-100; 148pp; English.  
 CC The present sequence represents recombinant adenovirus  
 CC Ad:Pac-beta-Gal, which exemplifies the invention. The specification  
 CC describes replication deficient adenoviruses having deletions of  
 CC the Ela and E3 regions and comprise a single packaging signal sequence  
 CC and Ela enhancer sequence, where the sequences are at the 3' end of the  
 CC adenovirus and the Ela deletion contains at least one PacI, ClaI, XbaI  
 CC or BstBI cleavage site. The replication deficient viral vectors can be  
 CC used in gene therapy regimens to effect the transfer of genes encoding  
 CC molecules of therapeutic importance, including isoforms of the nitric  
 CC oxide synthetase (NOS) gene (brain, endothelial and microphage NOS), the  
 CC cystic fibrosis chloride channel (CFTR) gene, the dystrophin gene, the  
 CC LDL receptor gene and the erythropoietin gene. The NOS isoforms can be  
 CC used in vascular applications or in cancer therapy (microphage NOS). The  
 CC NOS gene can be introduced into vein grafts prior to their use as  
 CC coronary artery bypass grafts. A NOS containing adenovirus can also be  
 CC used following coronary angioplasty to prevent rethrombosis and to treat  
 CC atherosclerotic arteries.  
 CC Sequence 34382 BP; 7923 A; 9880 C; 9421 G; 7158 T;

## Query Match

Best Local Similarity 77.0%; Score 15.4; DB 1; Length 34382;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19  
 ||||| |||||

Db 12336 CGCATGGCTGAATCG 12352

## RESULT 7

X07371  
 ID X07371 standard; DNA; 34427 BP.  
 AC X07371;  
 DT 07-JUN-1999 (first entry)  
 DE Adenovirus vector 231-10 genome.  
 KW Adenovirus; vector 231-10; apoptosis; inhibitor; RID-alpha;  
 KW RID-beta; receptor internalisation and degradation; cancer;

degenerative disease; immune disorder; gene therapy;  
tissue transplant; death receptor; tumour necrosis factor receptor;  
human; ds.  
Chimeric - Mastadenovirus 5.  
Chimeric - Mastadenovirus 2.  
Chimeric - Human cytomegalovirus.  
WO9902658-A1.  
21-JAN-1999.  
08-JUL-1998; U14239.  
09-JUL-1997; US-088993.  
(UYSL-) UNIV SAINT LOUIS.  
Wold WSM;  
WPI; 99-120862/10.  
Method for inhibiting apoptosis - using receptor internalisation and  
degradation (RID) complex  
Example 10; Fig. 28A-XX; 126pp; English.  
This is the nucleotide sequence of vector 231-10, a human  
adenovirus serotype 5 (Ad5) vector. 231-10 lacks the E1A, E1B  
and E3 transcription units. The deleted E1A and E1B regions are  
replaced with an expression cassette wherein all the E3 proteins  
are expressed from the human cytomegalovirus (CMV) promoter. The  
E3 expression cassette contains the E3 genes from virus pm34.1,  
a derivative of virus rec700, which is an Ad5-Ad2-Ad recombinant  
that has the Ad2 version of the E3 genes for the 12.5k, 6.7k,  
gp19K and RID-alpha proteins, and the Ad5 version of the E3 genes  
for the RID-beta and 14.7K proteins. In addition, the vector has  
2 missense mutations in the adp (adenovirus death protein) gene  
that eliminate the first 2 Met codons, thereby precluding synthesis  
of functional ADP. Because 231-10 lacks E1A, viral genes in the  
vector backbone are not expressed; only the E3 proteins are  
expressed from the CMV promoter. Thus, the vector serves as an  
essentially inert vehicle by which E3 genes can be delivered into  
cells. The invention provides a method for inhibiting apoptosis of  
a cell expressing a death receptor of the tumour necrosis factor  
receptor family. This involves treating the cells with a receptor  
internalisation and degradation (RID) protein complex containing  
RID-alpha (10.4 k, see W97877 and W97878) and RID-beta (14.5k, see  
W97879) proteins encoded by the E3 region of adenovirus. The cell  
can be treated by administering to the cell a polynucleotide  
expressing the RID complex or a composition containing the RID  
complex. The compositions and methods are used to treat  
degenerative immunodeficiency diseases, particularly to decrease  
leukocyte apoptosis (claimed). Immune disorders that can be  
treated include autoimmune disorders, ischaemic injury such as  
caused by myocardial infarction, stroke induced neuron death and  
reperfusion injury, alcohol-induced hepatitis, diseases caused by  
viral infection such as AIDS and fulminant hepatitis, and cancer.  
The methods can also be used to promote tissue transplant  
survival. Vector 231-10 can be used to deliver the RID complex to  
the cells.  
Sequence 34427 BP; 8098 A; 9606 C; 9451 G; 7272 T;  
Query Match 77.0%; Score 15.4; DB 1; Length 34427;  
Best Local Similarity 94.1%; Pred. No. 39;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3 cggcatggactgaatcg 19  
||||||| |||||||  
Db 12578 CGGCATGGCGCTGAATCG 12594  
RESULT 8  
T60557  
ID T60557 standard; DNA; 35000 BP.  
AC T60557;  
DT 05-MAR-1998 (first entry)  
DE Recombinant cis-acting adenovirus H5.020TKneo-int(LTR).  
KW Adenovirus H5.020TKneo(LTR); plasmid pADMLVneo-int;  
KW retrovirus; retrotransposition; transgene;  
KW gene therapy; vector; neomycin resistance; neo gene; ss.  
OS Chimeric - Mastadenovirus serotype 5.  
OS Chimeric - Moloney murine leukaemia virus.

OS Chimeric - Rattus sp.  
OS Chimeric - Synthetic.  
PN WO9715679-A1.  
PD 01-MAY-1997.  
PF 24-OCT-1996; U17176.  
PR 27-OCT-1995; US-005942.  
PA (UYPE-) UNIV PENNSYLVANIA.  
PI Kelley WM, Wilson JM;  
DR WPI; 97-259031/23.  
PT Recombinant replication defective virus - inserts transgene into  
PT host cell chromatin in the presence of transposase, providing stable  
PT and durable transgene expression  
PS Example 2; Fig 8A-I; 74pp; English.  
CC This is the DNA sequence of the cis-acting recombinant adenovirus  
CC H5.020TKneo-int(LTR). It was obtained by co-transfecting HEK293  
CC (ATCC CRL1573) cells with linearised plasmid pADMLVneo-int  
CC (see T60554) and Ad5dl7001. A trans-acting adenovirus (see  
CC T60558) was also produced, and retrotransposition and integration  
CC into host cells was demonstrated in HeLa cells infected with both  
CC viruses. Claimed novel recombinant replication defective viruses  
CC (A) comprise: (a) DNA of, or corresponding to, at least part of the  
CC viral genome, able to infect a mammalian cell; and (b) a first  
CC expression sequence (ES1) comprising a human gene (I) linked to  
CC regulatory sequences for expression, with the cassette flanked by  
CC the cis-acting terminal repeat sequences of a transposon, and ES1  
CC flanked by DNA of (a); (A) can infect a mammalian cell and can  
CC express (I) and transfer it to the cellular chromatin in vivo or in  
CC vitro in the presence of a transposase. Also new are: (1) a  
CC recombinant replication defective virus (B) containing (a) as above  
CC and a second expression sequence (ES2) containing a trans-acting  
CC transposase gene (II) plus regulatory sequences, flanked by DNA of  
CC (a) and able to express transposase in mammalian cells; and (2)  
CC mammalian cells which stably express (I) integrated into its  
CC chromatin, produced by infection with (A). (A) and (B) are used  
CC for delivery and stable integration of (I) into a host cell  
CC (claimed), e.g. in somatic gene therapy of genetic defects or  
CC deficiencies such as cystic fibrosis. (A) are also used to prepare  
CC recombinant retroviruses (RRV) (claimed). Because (I) becomes  
CC stably integrated, it provides longer lasting expression than genes  
CC introduced with conventional adenoviral vectors and the need for  
CC repeated administration is avoided. The transgene is inherited by  
CC progeny cells and the viral component is gradually degraded by the  
CC cell. When used for RRV production, (A) increase the ratio of RRV  
CC to empty retrovirus.  
SQ Sequence 35000 BP; 7810 A; 9673 C; 9271 G; 7046 T;  
Query Match 77.0%; Score 15.4; DB 1; Length 35000;  
Best Local Similarity 94.1%; Pred. No. 39;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3 cggcatggactgaatcg 19  
||||||| |||||||  
Db 13495 CGGCATGGCGCTGAATCG 13511  
RESULT 9  
T59272  
ID T59272 standard; DNA; 35408 BP.  
AC T59272;  
DT 27-AUG-1997 (first entry)  
DE Recombinant adenovirus H5.001CBLacZ.  
KW Minigene; human adenovirus type 5; open reading frame; promoter; in vivo;  
KW mouse mammary tumour virus; growth hormone; terminator; recombinant;  
KW packaging cell line; adeno-associated virus; ex vivo; gene therapy;  
KW inherited disease; cancer; genetic dysfunction; cystic fibrosis; ds.  
OS Synthetic.  
PH Key Location/Qualifiers  
FT misc\_feature 1..330 a  
FT /\*tag= "sequence derived from plasmid pAd.CBLacZ";  
FT /note= "includes sequences from Ad map units 0-1"  
FT promoter 370..928



```

CC introduced with conventional adenoviral vectors and the need for
CC repeated administration is avoided. The transgene is inherited by
CC progeny cells and the viral component is gradually degraded by the
CC cell. When used for RVV production, (A) increase the ratio of RVV
CC to empty retrovirus.
SO Sequence 36538 BP; 8621 A; 10490 C; 9909 G; 7518 T;

Query Match 77.0%; Score 15.4; DB 1; Length 36538;
Best Local Similarity 94.1%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 cggcatggactgaatcg 19
    ||||| |||||
Db 15033 CGGCATGGCCTGAATCG 15049

RESULT 12
ID Q33008 standard; DNA; 2220 BP.
AC Q33008;
DT 18-MAY-1993 (first entry)
DE Novel intestinal oncofetal gene.
KW oncofetal; onco-fetal; onco-fetal; intestine; gastrointestinal;
KW cancer; tumour; growth; benign; malignant; neoplasm; metastasis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 115..1908
   /*tag= a
PN US5171850-A.
PD 15-DEC-1992.
PF 31-AUG-1988; 239084.
PR 31-AUG-1988; US-239084.
PA (ONTA-) ONTARIO CANCER INST.
PI Buick RN, Filmus JE;
DR WPI; 93-008640/01.
PT DNA or RNA molecule encoding intestinal onco-fetal gene - useful
PT for detecting neoplastic cells, esp. intestinally-originating
PT prim. and metastatic tumours
PS Claim 1; Fig 8; 12pp; English.
CC This sequence represents a novel intestinal oncofetal gene. A cDNA
CC library was made from ploy(A)+ RNA obtd. from IEC-18 rat foetal
CC intestine cells using the lambda gt10 cloning system. 20,000
CC plaques were transferred onto duplicate nitrocellulose filters.
CC Plaques that were positive when probed with IEC-18 subtracted cDNA
CC and negative with NIK-52E (normal rat kidney) cDNA were selected for
CC secondary screening. Five of these plaques showed similar results
CC in the secondary screening. DNA was prepd. from the 5 plaques and
CC the cloned inserts isolated. All of them had the same size (2.2kb)
CC and when labelled hybridised with IEC-18 cells and fetal rat
CC intestine but not with other rat cell lines or adult rat intestine.
CC The insert from one lambda gt 10 clone was subsequently cloned into
CC the EcoRI site of pUC13 and named OCI-5.
SQ Sequence 2220 BP; 588 A; 531 C; 533 G; 568 T;

Query Match 74.0%; Score 14.9; DB 1; Length 2220;
Best Local Similarity 88.9%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 aacggcatggactgaatc 18
    ||||| |||||
Db 1033 AACGGCATGTACAGATC 1050

RESULT 13
ID N50226 standard; DNA; 24593 BP.
AC N50226;
DT 16-OCT-1991 (first entry)
DE Sequence of opine synthase gene.
KW Plant vector; Ti plasmid; T-DNA; ss.

```

```

OS Agrobacterium tumefaciens Ti plasmid.
PN EP-140556-A.
PD 08-MAY-1985.
PF 14-SEP-1984; 193841.
PR 14-SEP-1983; US-532280.
PA (AGRI-) AGRIGENETICS RES.
PA (LUBR ) LUBRIZOL GENETICS.
PI Dahl GA, Sutton DW, Barker RF;
DR WPI; 85-112088/19.
PT Plasmid contg. opine synthase gene for selection - and foreign
PT DNA, useful as vector for transforming plant cells
PS Disclosure; Page 212-217; 69pp; Japanese.
CC Plant cells (and protoplasts) and plasmids contg. the DNA fragment
CC which includes an opine synthase gene plus a gene for antibiotic
CC resistance are claimed. These plasmids provide max. efficiency for
CC transfer of foreign genes and can be amplified in the plant genome.
CC They do not contain genes specifying tumour formation and will not
CC spread antibiotic resistance throughout the plant population.
SQ Sequence 24593 BP; 6554 A; 5493 C; 5787 G; 6759 T;

Query Match 72.0%; Score 14.4; DB 1; Length 24593;
Best Local Similarity 93.8%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 ggcattggactgaatcg 19
    ||||| |||||
Db 18201 GGCATGTACTGAATCG 18216

RESULT 14
ID N50182 standard; DNA; 24596 BP.
AC N50182;
DT 17-OCT-1991 (first entry)
DE Complete nucleotide sequence of the T-DNA region of the octopine Ti
DE plasmid pTi15955.
KW Plant vector; transformation-inducing principle (TIP) gene;
KW octopine Ti plasmid; ss.
OS Agrobacterium tumefaciens ATCC 15955.
PN EP-145338-A.
PD 19-JUN-1985.
PF 16-NOV-1984; 307969.
PR 18-NOV-1983; US-553786
PA (AGRI-) AGRIGENETICS RES.
PA (LUBR ) LUBRIZOL GENETICS.
PI Barker RF, Kemp JD;
DR WPI; 85-148223/25.
PT New DNA vectors contg. T-DNA sequence of octopine Ti plasmid -
PT for expression in plant cells to confer desirable properties to
PT plants and their cells
PS Claim 28; Fig 1; 87pp; English.
CC The inventors claim a vector contg. a transformation-inducing
CC principle (TIP) gene from Ti plasmid pTi15955. The sequence of the
CC T-DNA of the octopine-type Ti plasmid has fourteen open reading
CC frames bounded by eukaryotic promoters, ribosome binding sites, and
CC polyadenylation sites. With the vectors, expression of structural
CC foreign genes in plant cells is promoted. The gene esp. encodes an
CC insecticidal toxin identical to or derived from the crystal protein
CC of Bacillus thuringiensis.
SQ Sequence 24596 BP; 6534 A; 5510 C; 5793 G; 6759 T;

Query Match 72.0%; Score 14.4; DB 1; Length 24596;
Best Local Similarity 93.8%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 ggcattggactgaatcg 19
    ||||| |||||
Db 18203 GGCATGTACTGAATCG 18218

RESULT 15

```

T67032  
ID T67032 standard; DNA; 483 BP.  
AC T67032;  
DT 04-AUG-1997 (first entry)  
DE Salmonella iagB gene.  
KW iagB; Salmonella secreted protein; Ssp;  
KW bacterial-mediated endocytosis; diagnosis; therapy; vaccine;  
KW attenuation; virulence; ss.  
OS Salmonella typhimurium.  
FH Key Location/Qualifiers  
FT cds 1..483  
FT /\*tag= a  
FT /product= iagB  
FT /note= "in-frame stop codons at bases 136-138,  
FT 187-189, 211-213 and 220-222"  
PN W09718225-A1.  
PD 22-MAY-1997.  
PF 14-NOV-1996; U18504.  
PR 14-NOV-1995; US-006733.  
PA (GEHO ) GEN HOSPITAL CORP.  
PI Miller SI:  
DR WPI: 97-289217/26.  
DR P-PSDB; W15272.  
PT New isolated Salmonella secreted proteins and related genes - used  
PT to develop products for the detection, treatment or prevention of  
PT Salmonella infections  
PS Disclosure; Fig 24: 95pp; English.  
CC A DNA sequence (T67032) codes for the iagB polypeptide (W15272)  
CC of Salmonella typhimurium. The invention features proteins  
CC (see also W15268-71) and nucleic acids (T67028-31) involved in  
CC Salmonella typhimurium virulence and/or bacterial-mediated  
CC endocytosis.  
SQ Sequence 483 BP; 159 A; 89 C; 110 G; 125 T;

Query Match 71.0%; Score 14.2; DB 1; Length 483;  
Best Local Similarity 84.2%; Pred. NO. 88;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 acggcatggactgaatcgg 20  
||||||| ||| |  
Db 46 ACGGCATGGGCTGATCTG 64

Search completed: May 23, 2000, 09:42:23  
Job time: 8855 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:21:48 ; Search time 1428.99 Seconds  
(without alignments)  
56.729 Million cell updates/sec

Title: US-08-945-805-4  
Perfect score: 20  
Sequence: 1 aacggcatgactgaatcgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*

- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_est45:\*
- 80: gb\_est46:\*
- 81: gb\_est47:\*
- 82: gb\_gss1:\*
- 83: gb\_gss2:\*
- 84: gb\_gss3:\*
- 85: gb\_gss4:\*
- 86: em\_gss1:\*
- 87: em\_gss2:\*
- 88: em\_gss3:\*
- 89: em\_gss4:\*
- 90: gb\_gss5:\*
- 91: gb\_gss6:\*
- 92: gb\_gss7:\*
- 93: gb\_gss8:\*
- 94: gb\_gss9:\*
- 95: em\_gss5:\*
- 96: em\_gss6:\*
- 97: em\_gss7:\*
- 98: em\_gss8:\*
- 99: em\_gss9:\*
- 100: em\_gss10:\*
- 101: em\_gss11:\*
- 102: gb\_gss10:\*
- 103: gb\_gss11:\*
- 104: em\_gss12:\*
- 105: gb\_gss12:\*
- 106: gb\_gss13:\*
- 107: gb\_gss14:\*
- 108: gb\_gss15:\*
- 109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result      %      Query

No.	Score	Match	Length	DB	ID	Description
1	17.4	87.0	743	102	AQ157804	AQ157804 nbxb00101r
2	16.8	84.0	514	85	AQ726116	AQ726116 HS_5403_B
3	16.8	84.0	517	85	AQ692319	AQ692319 HS_5408_A
4	16	80.0	562	105	AQ299540	AQ299540 HS_3021_A
5	15.8	79.0	327	62	AI904349	AI904349 IL-BT051-
6	15.8	79.0	426	84	B41684	B41684 HS-1054-A1-
7	15.8	79.0	496	108	AQ570714	AQ570714 HS_5355_B
8	15.8	79.0	504	31	AA290080	AA290080 vc93d10.r
9	15.8	79.0	568	85	AQ717770	AQ717770 HS_5507_B
10	15.8	79.0	670	47	AI487266	AI487266 EST245588
11	15.4	77.0	144	27	C01417	C01417 HUMGS000840
12	15.4	77.0	163	24	H61188	H61188 yu40d04.s1
13	15.4	77.0	187	34	AA496685	AA496685 zv38g09.s
14	15.4	77.0	186	49	AV026011	AV026011 AV026011
15	15.4	77.0	201	73	AV357777	AV357777 AV357777
16	15.4	77.0	202	24	H62092	H62092 yu40d04.r1
17	15.4	77.0	212	34	AA481830	AA481830 zv38g09.r
18	15.4	77.0	214	50	AI672598	AI672598 we55b10.x
19	15.4	77.0	242	71	AV316747	AV316747 AV316747
20	15.4	77.0	242	71	AV320720	AV320720 AV320720
21	15.4	77.0	254	60	AV145419	AV145419 AV145419
22	15.4	77.0	264	70	AV220334	AV220334 AV220334
23	15.4	77.0	279	59	AI766512	AI766512 w141e11.x
24	15.4	77.0	282	70	AV228365	AV228365 AV228365
25	15.4	77.0	288	47	AI501322	AI501322 UI-R-C2p-
26	15.4	77.0	291	39	AA25301	AA25301 cc67d10.s
27	15.4	77.0	293	49	AI657048	AI657048 tt55601.x
28	15.4	77.0	299	70	AV224977	AV224977 AV224977
29	15.4	77.0	307	46	AA999126	AA999126 UI-R-E1-f
30	15.4	77.0	340	29	AA185899	AA185899 mt91b10.r
31	15.4	77.0	340	35	AA581745	AA581745 nn48a02.r
32	15.4	77.0	352	38	AA745863	AA745863 ny93d01.s
33	15.4	77.0	373	41	AI032406	AI032406 cv37b09.x
34	15.4	77.0	374	74	AW204826	AW204826 UI-H-B11-
35	15.4	77.0	375	105	AQ295180	AQ295180 HS_3065_A
36	15.4	77.0	391	50	AI682721	AI682721 wc65c12.x
37	15.4	77.0	392	46	AA964313	AA964313 UI-R-C0-g
38	15.4	77.0	395	35	AA579662	AA579662 nm72b07.s
39	15.4	77.0	395	46	AA997899	AA997899 UI-R-C0-h
40	15.4	77.0	399	45	AI011808	AI011808 EST206259
41	15.4	77.0	411	61	AI852103	AI852103 UI-M-BH0-
42	15.4	77.0	415	41	AI024518	AI024518 ou51b04.x
43	15.4	77.0	419	49	AI628407	AI628407 ey76b04.x
44	15.4	77.0	422	49	AI657058	AI657058 tt55f01.x
45	15.4	77.0	430	47	AI499687	AI499687 tm91c12.x

## ALIGNMENTS

RESULT 1  
 AQ157804 743 bp DNA GSS 12-SEP-1998  
 LOCUS nbxb0010117r CUGI Rice BAC Library Oryza sativa genomic clone  
 DEFINITION nbxb0010117r, genomic survey sequence.

ACCESSION AQ157804  
 VERSION AQ157804.1 GI:3554829  
 KEYWORDS GSS.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
 1 (bases 1 to 743)  
 WING, R.A., and Dean, R.A.  
 A BAC End Sequencing Framework to Sequence the Rice Genome  
 Unpublished (1998)  
 Contact: Wing RA  
 Clemson University  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: GGAAACAGCTATGACCATG  
 Class: BAC ends  
 High quality sequence stop: 346.  
 Location/Qualifiers  
 1..743  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbxb0010117r"  
 /clone.lib="CUGI Rice BAC Library"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /notes="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

FEATURES  
source

BASE COUNT 201 a 153 c 148 g 241 t  
 ORIGIN

Query Match 87.0%; Score 17.4; DB 102; Length 743;  
 Best Local Similarity 94.7%; Pred. No. 37;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 accgatggactgaatcgg 20  
 |||||  
 Db 662 ACGCATGGACCGAATCG 680

RESULT 2  
AQ726116/c

LOCUS AQ726116 514 bp DNA GSS 14-JUL-1999  
 DEFINITION HS\_5403\_B2\_E11\_T7A RPCI-II Human Male BAC Library Homo sapiens genomic clone Plate=979 Col=22 Row=J, genomic survey sequence.  
 ACCESSION AQ726116  
 VERSION AQ726116.1 GI:5485785  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 514)  
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., and Kaller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 REFERENCE Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 99380589  
 JOURNAL MEDLINE  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618

BASE COUNT		162 a	110 c	79 g	151 t	15 others
ORIGIN						
<hr/>						
Query Match 84.0%; Score 16.8; DB 85; Length 517;						
Best Local Similarity 90.0%; Pred. No. 71;						
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;						
<hr/>						
QY	1	aacggcgtgactgaatcg	20			
Db	385	AATGGCATAGACTGAATCGG	366			
<hr/>						
RESULT	4					
LOCUS AQ299540/c						
DEFINITION HS_3021.AL_B09.T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3021 Col=17 Row=C, genomic survey sequence.						
ACCESSION	AQ299540	DNA	GSS	15-DEC-1998		
VERSION	AQ299540.1	GI:4016719				
KEYWORDS	GSS.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 562)					
	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.					
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome					
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)					
MEDLINE	99380589					
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3021 row: C column: 17 Class: BAC ends High quality sequence stop: 562. Location/Qualifiers 1..562					
<hr/>						
FEATURES		source				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/clone="plate=3021 Col=17 Row=C"				
		/clone_lib="CIT Approved Human Genomic Sperm Library D"				
		/sex="male"				
		/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"				
BASE COUNT		125 a	156 c	74 g	201 t	6 others
ORIGIN						
<hr/>						
Query Match 80.0%; Score 16; DB 105; Length 562;						
Best Local Similarity 100.0%; Pred. No. 1.9e+02;						
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
<hr/>						
QY	1	aacggcgtgactgaa	16			
Db	210	AACGGCATGGACTGAA	195			

```

RESULT 5
AI904349/c      327 bp      mRNA      EST      01-DEC-1999
LOCUS          IL-BT051-011298-007 BT051 Homo sapiens cDNA, mRNA sequence.
DEFINITION     AI904349
ACCESSION      AI904349
VERSION        AI904349.1 GI:6494736
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 327)
AUTHORS        HCGP http://www.ludwig.org.br/ORESTES.
TITLE          The FAPESP/LICR Human Cancer Genome Project
JOURNAL        Unpublished (1999)
COMMENT        On May 9, 1996 this sequence version replaced gi:1301793.
               Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/seq/gethtml.pl?tl=tl&t2=tl-BT051-007.html&t3=011298&t4=1)
               Seq primer: puc 18 forward.
FEATURES       Location/Qualifiers
               source          1..327
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /clone_lib="BT051"
                           /sex="female"
                           /dev_stage="Adult"
                           /note="Organ: breast; Vector: puc18; Site:1: SmaI; Site:2:
                           SmaI; A mini-library was made by cloning products derived
                           from ORESTES PCR (U.S. Letters Patent application No.
                           190,716 - Ludwig Institute for Cancer Research) profiles
                           into the puc 18 vector. Reverse transcription of tissue
                           mRNA and cDNA amplification were performed under low
                           stringency conditions."
BASE COUNT     60 a 113 c 25 g 119 t 10 others
ORIGIN
1 acgcgctggactgaatcg 19
||||| |||||||||
Db 298 AACGGAATGGACTGAATCG 280

RESULT 6
B41684
LOCUS          HS-1054-A1-H01-MR.abi CIT Human Genomic Sperm Library C Homo
DEFINITION     sapiens genomic clone Plate=Ct 776 Col=1 Row=O, genomic survey
               sequence.
ACCESSION      B41684
VERSION        B41684.1 GI:2545936
KEYWORDS       GSS.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 426)
AUTHORS        Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., SchmidT,S.,
               Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
TITLE          Construction of a Characterized Clone Resource for Genomic

Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
Tagged Connectors
Unpublished (1997)
Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu
Sequence Tagged Connector
Plate: CT 776 row: O column: 1
Class: BAC ends
High quality sequence stop: 426.
FEATURES       Location/Qualifiers
               source          1..426
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /clone_lib="Plate=Ct 776 Col=1 Row=O"
                           /clone_lib="CIT Human Genomic Sperm Library C"
                           /sex="M"
                           /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
                           E-Coli DH10B"
BASE COUNT     145 a 43 c 98 g 138 t 2 others
ORIGIN
1 acggcatggactgaatcg 20
||||| |||||||
Db 132 ACGCATGGCATGAATCG 150

Query Match      79.0%; Score 15.8; DB 84; Length 426;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 acggcatggactgaatcg 20
||||| |||||||
Db 132 ACGCATGGCATGAATCG 150

RESULT 7
AQ570714/c
LOCUS          AQ570714 496 bp      DNA      GSS      01-JUN-1999
DEFINITION     HS-5355_B1_D12_SPOE RPCI-11 Human Male BAC Library Homo sapiens
               genomic clone Plate=931 Col=23 Row=H, genomic survey sequence.
ACCESSION      AQ570714
VERSION        AQ570714.1 GI:4963934
KEYWORDS       GSS.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 496)
AUTHORS        Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
               Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
               Hood,L.
TITLE          Sequence-tagged connectors: A sequence approach to mapping and
               scanning the human genome
JOURNAL        Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE        99380589
COMMENT        On Feb 19, 1999 this sequence version replaced gi:4129325.
               Contact: Mahairas GG, Wallace JC, Hood L
               High Throughput Sequencing Center
               University of Washington
               401 Queen Anne Avenue North, Seattle, WA 98109, USA
               Tel: (206) 616-3618
               Fax: (206) 616-3887
               Email: jwallace@u.washington.edu
               Clones are derived from the human BAC library RPCI-11. For BAC
               library availability, please contact Pieter de Jong
               (pieter@dejong.med.buffalo.edu). Clones may be purchased from
               BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)
               or from Resear h Genetics (http://bacpac.med.buffalo.edu/ordering\_bac.htm)
               http://www.htsc.washington.edu
               Plate: 931 row: H column: 23
               Seq primer: SP6
               Class: BAC ends
               High quality sequence stop: 496.

```

```
FEATURES
source
  Location/Qualifiers
    1. 496
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="Plate-931 Col=23 Row=H"
      /sex="male"
      /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      100 a   146 c   89 g   160 t   1 others
ORIGIN
Query Match      79.0%; Score 15.8; DB 108; Length 496;
Best Local Similarity 89.5%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 aacggcgtgactgaatcg 19
    ||||| ||||| |||||
Db 170 AACGGCATGAATGAACG 152

RESULT 8
AA290080
LOCUS      AA290080      504 bp      mRNA      EST      14-APR-1997
DEFINITION vc93d10.r1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:790579 5',
            mRNA sequence.
ACCESSION  AA290080
VERSION    AA290080.1 GI:1936439
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 504)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE      The WashU-HMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    On Sep 12, 1996 this sequence version replaced gi:1392826.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:474987
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 489.
FEATURES
source
  Location/Qualifiers
    1. 504
      /organism="Mus musculus"
      /strain="BALB/c"
      /db_xref="taxon:10090"
      /clone="IMAGE:790579"
      /clone_lib="Barstead MPLRB1"
      /sex="mixed"
      /tissue_type="kidney"
      /dev_stage="6 weeks"
      /lab_host="DH10B"
      /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAACTGAAGTGGAGCGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
            3']; double-stranded cDNA was ligated to Eco RI adaptors
            [CATGATTCGGTACC], digested with Not I and cloned into the
            Not I and Eco RI sites of the modified pT73 vector.
            Library constructed by Bob Barstead."
```

```
BASE COUNT      139 a   120 c   157 g   88 t
ORIGIN
Query Match      79.0%; Score 15.8; DB 31; Length 504;
Best Local Similarity 89.5%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 aacggcgtgactgaatcg 19
    ||||| ||||| ||||| |||||
Db 86 AACGGCCAGGACTGAATCG 104

RESULT 9
AQ717770/c
LOCUS      AQ717770      568 bp      DNA      GSS      13-JUL-1999
DEFINITION HS_5507_B1_D05_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1083 Col=9 Row=H, genomic survey sequence.
ACCESSION  AQ717770
VERSION    AQ717770.1 GI:5467086
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 568)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            plate: 1083 row: H column: 9
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 568.
FEATURES
source
  Location/Qualifiers
    1. 568
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="Plate=1083 Col=9 Row=H"
      /clone_lib="RPCI-11 Human Male BAC Library"
      /sex="male"
      /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      165 a   133 c   109 g   145 t   16 others
ORIGIN
Query Match      79.0%; Score 15.8; DB 85; Length 568;
Best Local Similarity 89.5%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 acggcgtgactgaatcg 20
    ||||| ||||| ||||| |||||
Db 97 ACGGCATGTACTGCATCG 79

RESULT 10
AT487266/c
```



```

/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bernaldo."
BASE COUNT      41 a      36 c      54 g      32 t
ORIGIN

Query Match      77.0%; Score 15.4; DB 24; Length 163;
Best Local Similarity 94.1%; Pred. No. 3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggcattggactgaatcgg 20
|||||
Db 65 GCATGGAGTGAATCTG 49

RESULT 13
LOCUS AA496685/c 187 bp mRNA EST 12-AUG-1997
DEFINITION zv38g09.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
IMAGE:755968 3', mRNA sequence.
ACCESSION AA496685
VERSION AA496685.1 GI:2230006
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Kucaba, T., Lucy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
TITLE Unpublished (1997)
JOURNAL On Jan 25, 1995 this sequence version replaced gi:637862.
COMMENT Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amer-shan.
FEATURES
source
Location/Qualifiers
1..187
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:755968"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bernaldo."

```

```

BASE COUNT      46 a      40 c      59 g      42 t
ORIGIN

Query Match      77.0%; Score 15.4; DB 34; Length 187;
Best Local Similarity 94.1%; Pred. No. 3.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggcattggactgaatcgg 20
|||||
Db 88 GCATGGAGTGAATCTG 72

RESULT 14
LOCUS AV026011 196 bp mRNA EST 31-AUG-1999
DEFINITION AV026011 Mus musculus adult C57BL/6J lung Mus musculus cDNA clone
1200017F11, mRNA sequence.
ACCESSION AV026011
VERSION AV026011.1 GI:4803003
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 196)
AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Tateno, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
TITLE Unpublished (1999)
JOURNAL On Apr 7, 1998 this sequence version replaced gi:3034525.
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermolabile and its application for the synthesis of full length cDNA
trehalose and its application for the synthesis of full length cDNA
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
Location/Qualifiers
source
1..196
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1200017F11"
/clone_lib="Mus musculus adult C57BL/6J lung"
/tissue_type="lung"
/dev_stage="adult"
BASE COUNT      45 a      48 c      38 g      65 t
ORIGIN

Query Match      77.0%; Score 15.4; DB 49; Length 196;
Best Local Similarity 94.1%; Pred. No. 3.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 aacggcattggactgaat 17
|||||
Db 100 AAGGCGATGGACTGAAT 116

```

RESULT 15  
 AV357777/c  
 LOCUS  
 DEFINITION AV357777 RIKEN full-length enriched, in vitro fertilized eggs Mus  
 musculus cDNA clone 7420409B05 3', mRNA sequence.  
 ACCESSION AV357777  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM house mouse.  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 201)  
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,  
 Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,  
 Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I.,  
 Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,  
 Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
 Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N.,  
 Yokota, T., Yoshiki, A., Washinabe, S., Yamamura, T., Yasunishi, A.,  
 RIKEN Mouse ESTs (Konno, H., et al.)  
 Unpublished (1999)  
 On Jul 9, 1999 this sequence version replaced gi:5435173.  
 Contact: Yoshihide Hayashizaki  
 Genome Exploration Research Group, Life Science Tsukuba Center,  
 The Institute of Physical and Chemical Research (RIKEN), Genomic  
 Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: +81-298-36-9013  
 Fax: +81-298-36-9098  
 Email: genome-resetc.riken.go.jp/  
 URL: http://genome.rtc.riken.go.jp/  
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
 Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and  
 Hayashizaki, Y.  
 Transcriptional sequencing: A method for DNA sequencing using RNA  
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitsumi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
 Okazaki, Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

# TITLE JOURNAL COMMENT

## FEATURES source

Location/Qualifiers  
 1. .201  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="7420409B05"  
 /clone\_lib="RIKEN full-length enriched, in vitro  
 fertilized eggs"  
 /sex="female"  
 /tissue\_type="in vitro fertilized eggs"  
 /dev\_stage="egg"  
 /lab\_host="Dhl08"  
 /note="Site\_1: Sali; Site\_2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5'  
 GAGAGAGAGATTCGAGTTAATAATCCCCCCCCCCC 3']. cDNA  
 was cloned into the XhoI and BamHI sites. Vector: a  
 modified pBluescript KS(+) after bulk excision from Lambda  
 FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI"

BASE COUNT 55 a 44 c 52 g 50 t  
 ORIGIN  
 Query Match 77.0%; Score 15.4; DB 73; Length 201;  
 Best Local Similarity 94.1%; Pred. No. 3.1e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 aacggcactggactgaat 17  
 ||||| ||||| ||||| |||||  
 Db 101 AACGGGATGGACTGAAT 85

Search completed: May 23, 2000, 09:21:53  
 Job time: 7780 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2000, 09:40:33 ; Search time 71.7 Seconds  
(without alignments)  
36.258 Million cell updates/sec

Title: US-08-945-805-4  
Perfect score: 20  
Sequence: 1 aacggcagtgactgaatcg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES										
Result No.	Score	Query Match	Length	ID	Description					
1	15.4	77.0	34303	4	US-08-735-609-4	Sequence 4, Appli				
2	15.4	77.0	34382	3	US-08-374-483-6	Sequence 6, Appli				
3	15.4	77.0	35935	4	US-08-735-609-1	Sequence 1, Appli				
4	15.4	77.0	35935	5	US-08-379-452-43	Sequence 43, Appli				
5	14.8	74.0	2220	7	5171850-1	Patent No. 5171850				
6	14.4	72.0	24595	7	5428147-1	Patent No. 5428147				
c 7	14.2	71.0	1635	6	PCT-US94-05795-1	Sequence 1, Appli				
8	14.2	71.0	1713	4	US-08-486-839-5	Sequence 5, Appli				
9	14.2	71.0	1894	1	US-08-476-008-66	Sequence 66, Appli				
10	14.2	71.0	1894	1	US-08-306-063-66	Sequence 66, Appli				
11	14.2	71.0	1894	2	US-08-833-485-66	Sequence 66, Appli				
12	13.8	69.0	22	3	US-08-531-070A-9	Sequence 9, Appli				
c 13	13.8	69.0	1756	1	US-07-879-647A-28	Sequence 28, Appli				
c 14	13.8	69.0	1756	1	US-07-879-647A-30	Sequence 30, Appli				
c 15	13.8	69.0	1756	1	US-07-879-584A-28	Sequence 28, Appli				
c 16	13.8	69.0	1756	1	US-07-879-584A-30	Sequence 30, Appli				
c 17	13.8	69.0	1756	1	US-07-879-470A-28	Sequence 28, Appli				
c 18	13.8	69.0	1756	1	US-07-879-470A-30	Sequence 30, Appli				
c 19	13.8	69.0	1756	1	US-07-879-644A-28	Sequence 28, Appli				
c 20	13.8	69.0	1756	1	US-07-879-644A-30	Sequence 30, Appli				
c 21	13.8	69.0	1756	1	US-07-879-640A-28	Sequence 28, Appli				
c 22	13.8	69.0	1756	1	US-07-879-640A-30	Sequence 30, Appli				
c 23	13.8	69.0	1756	1	US-07-879-594A-28	Sequence 28, Appli				
c 24	13.8	69.0	1756	1	US-07-879-594A-30	Sequence 30, Appli				
c 25	13.8	69.0	1756	1	US-07-879-469A-28	Sequence 28, Appli				
c 26	13.8	69.0	1756	1	US-07-879-469A-30	Sequence 30, Appli				
c 27	13.8	69.0	3853	5	US-09-193-107-1	Sequence 1, Appli				

c 28	13.6	68.0	508	1	US-07-879-647A-50	Sequence 50, Appli
c 29	13.6	68.0	508	1	US-07-879-584A-50	Sequence 50, Appli
c 30	13.6	68.0	508	1	US-07-879-470A-50	Sequence 50, Appli
c 31	13.6	68.0	508	1	US-07-879-644A-50	Sequence 50, Appli
c 32	13.6	68.0	508	1	US-07-879-640A-50	Sequence 50, Appli
c 33	13.6	68.0	508	1	US-07-879-594A-50	Sequence 50, Appli
c 34	13.6	68.0	508	1	US-07-879-469A-50	Sequence 50, Appli
c 35	13.6	68.0	958	3	US-08-632-514C-10	Sequence 10, Appli
c 36	13.6	68.0	1747	2	US-08-327-516-1	Sequence 1, Appli
c 37	13.6	68.0	2169	5	US-08-806-326-5	Sequence 5, Appli
c 38	13.6	68.0	3564	3	US-07-594-921C-9	Sequence 9, Appli
c 39	13.6	68.0	4837	1	US-08-089-986-1	Sequence 1, Appli
c 40	13.6	68.0	4837	1	US-08-478-585-1	Sequence 1, Appli
c 41	13.6	68.0	4837	2	US-08-717-312-1	Sequence 1, Appli
c 42	13.6	68.0	4837	3	US-08-266-408-1	Sequence 1, Appli
c 43	13.6	68.0	4837	6	PCT-US94-07886-1	Sequence 1, Appli
c 44	13.6	68.0	5561	3	US-08-400-159-1	Sequence 1, Appli
c 45	13.6	68.0	246240	3	US-08-724-394A-20	Sequence 20, Appli

ALIGNMENTS

RESULT 1  
US-08-735-609-4  
; Sequence 4, Application US/08735609  
; Patent No. 5955360  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; APPLICANT: Amalfitano, Andrea  
; APPLICANT: Hauser, Michael A.  
; APPLICANT: Kumar-Singh, Rajendra  
; APPLICANT: Hartigan-O'Connor, Dennis J.  
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08735,609  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40, 027  
; REFERENCE/DOCKET NUMBER: UM-02484  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 34303 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "DNA"  
US-08-735-609-4

Query Match 77.0%; Score 15.4; DB 4; Length 34303;  
Best Local Similarity 94.1%; Pred. No. 27;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 3 cggcatggactgaatcg 19
    ||||| |||||
Db 12914 CGGCATGGCCTGAATCG 12930

RESULT 2
US-08-374-483-6
; Sequence 6, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-6

Query Match 77.0%; Score 15.4; DB 3; Length 34382;
Best Local Similarity 94.1%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19
    ||||| |||||
Db 12336 CGGCATGGCCTGAATCG 12352

RESULT 3
US-08-735-609-1
; Sequence 1, Application US/08735609
; Patent No. 5955360
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco

QY 3 cggcatggactgaatcg 19
    ||||| |||||
Db 12914 CGGCATGGCCTGAATCG 12930

STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35935 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-735-609-1

Query Match 77.0%; Score 15.4; DB 4; Length 35935;
Best Local Similarity 94.1%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcatggactgaatcg 19
    ||||| |||||
Db 11413 CGGCATGGCCTGAATCG 11429

RESULT 4
US-08-379-452-43
; Sequence 43, Application US/08379452
; Patent No. 6040174
; GENERAL INFORMATION:
; APPLICANT: IMLER, Jean-Luc
; APPLICANT: MEHTALI, Majid
; APPLICANT: PAVIRANI, Andrea
; TITLE OF INVENTION: DEFECTIVE ADENOVIRUSES AND CORRESPONDING
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: 1737 King Street, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22314-2756
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,452
; FILING DATE: 26-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR94/00624
; FILING DATE: 27-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93 06482
; FILING DATE: 28-MAY-1993
; ATTORNEY/AGENT INFORMATION:
```

; NAME: Dadio, Susan M.  
; REGISTRATION NUMBER: 40,373  
; REFERENCE/DOCKET NUMBER: 029395-002  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 35935 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-379-452-43

Query Match 77.0%; Score 15.4; DB 5; Length 35935;  
Best Local Similarity 94.1%; Pred. No. 27;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cggcgtggaactgaatcg 19  
||||| |||||||  
Db 11413 CGGCATGGCGCTGAATCG 11429

RESULT 5  
5171850-1  
; Patent No. 5171850  
; APPLICANT: FILMUS, JORGE E.; BUICK, RONALD N.  
; TITLE OF INVENTION: INTESTINAL ONCOFETAL GENE  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/239,084  
; FILING DATE: 31-AUG-1988  
; SEQ ID NO: 1:  
; LENGTH: 2220  
5171850-1

Query Match 74.0%; Score 14.8; DB 7; Length 2220;  
Best Local Similarity 88.9%; Pred. No. 36;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacggcgtggaactgaatc 18  
||||| || |||||  
Db 1033 aacggcgtgtacagaatc 1050

RESULT 6  
5428147-1  
; Patent No. 5428147  
; APPLICANT: BARKER, RICHARD F.; KEMP, JOHN D.  
; TITLE OF INVENTION: OCTOPINE T-DNA PROMOTERS  
; NUMBER OF SEQUENCES: 17  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/91,538  
; FILING DATE: 13-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 869,216  
; FILING DATE: 13-APR-1992  
; APPLICATION NUMBER: 869,216  
; FILING DATE: 13-APR-1992  
; APPLICATION NUMBER: 440,432  
; FILING DATE: 21-NOV-1989  
; APPLICATION NUMBER: 553,786  
; FILING DATE: 19-NOV-1983  
; APPLICATION NUMBER: 741,034  
; FILING DATE: 06-AUG-1991  
; APPLICATION NUMBER: 144,775  
; FILING DATE: 20-JAN-1988  
; APPLICATION NUMBER: 485,614  
; FILING DATE: 15-APR-1983  
; APPLICATION NUMBER: 713,624  
; FILING DATE: 10-JUN-1991  
; APPLICATION NUMBER: 260,574  
; FILING DATE: 21-OCT-1988

; APPLICATION NUMBER: 848,733  
; FILING DATE: 01-APR-1986  
; APPLICATION NUMBER: 535,354  
; FILING DATE: 26-SEP-1983  
; SEQ ID NO: 1:  
; LENGTH: 24595  
5428147-1

Query Match 72.0%; Score 14.4; DB 7; Length 24595;  
Best Local Similarity 93.8%; Pred. No. 84;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggcatggactgaatcg 19  
||||| |||||||  
Db 18203 ggcatgtactgaatcg 18218

RESULT 7  
PCT-US94-05795-1/c  
; Sequence 1, Application PC/TUS9405795  
; GENERAL INFORMATION:  
; APPLICANT: GENENCOR INTERNATIONAL, INC.  
; TITLE OF INVENTION: Isolation of Mushroom-Inducing Genes  
; TITLE OF INVENTION: and  
; TITLE OF INVENTION: Their Use in DNA-Mediated Transformation of Edible  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genencor  
; STREET: 180 Kimball Way  
; CITY: South San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94080

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/05795  
; FILING DATE:

; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Krupen, Karen I.  
; REGISTRATION NUMBER: 34,647  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 742-7500  
; TELEFAX: (415) 742-7217  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1635 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
PCT-US94-05795-1

Query Match 71.0%; Score 14.2; DB 6; Length 1635;  
Best Local Similarity 84.2%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acggcgtggaactgaatcgg 20  
| || | |||||||  
Db 441 AAGGCTTGACTGAATCGG 423

RESULT 8  
US-08-486-839-5  
; Sequence 5, Application US/08486839  
; Patent No. 5928928

```
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: A human chitinase, its recombinant
; TITLE OF INVENTION: production, its use for decomposing chitin, its use
; TITLE OF INVENTION: in therapy or prophylaxis against infection diseases.
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11758
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,839
; FILING DATE: 07 - June - 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baron, Ronald J.
; REGISTRATION NUMBER: 29,281
; REFERENCE/DOCKET NUMBER: 294-26
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
;
US-08-486-839-5

Query Match 71.0%; Score 14.2; DB 4; Length 1713;
Best Local Similarity 84.2%; Pred. No. 71;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 acggcagtgactgaatcgg 20
Db 1155 ACGCAGGACTGAATGGG 1173

RESULT 9
US-08-476-008-66
; Sequence 66, Application US/08476008
; Patent No. 5627061
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: Glyphosate Tolerant
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
; STREET: 700 Chesterfield Village Parkway
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,063
; FILING DATE: 13-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,611
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,008
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,063
; FILING DATE: 13-SEP-1994
; APPLICATION NUMBER: US 07/749,611
; FILING DATE: 28-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/576,537
; FILING DATE: 31-AUG-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoerner Jr., Dennis R.
; REGISTRATION NUMBER: 30,914
; REFERENCE/DOCKET NUMBER: 38-21(10660)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-6099
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 275..1618
;
US-08-476-008-66

Query Match 71.0%; Score 14.2; DB 1; Length 1894;
Best Local Similarity 84.2%; Pred. No. 72;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aacggcagtgactgaatcg 19
Db 1720 ATCGGCATGGACCGTATCG 1738

RESULT 10
US-08-306-063-66
; Sequence 66, Application US/08306063
; Patent No. 5633435
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: Glyphosate Tolerant
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F
; STREET: 700 Chesterfield Village Parkway
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,063
; FILING DATE: 13-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,611
```

; FILING DATE: 28-AUG-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/576,537  
; FILING DATE: 31-AUG-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoerner Jr., Dennis R.  
; REGISTRATION NUMBER: 30,914  
; REFERENCE/DOCKET NUMBER: 38-21(10660)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)537-6099  
; TELEFAX: (314)537-6047  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1894 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 275..1618  
US-08-306-063-66

Query Match 71.0%; Score 14.2; DB 1; Length 1894;  
Best Local Similarity 84.2%; Pred. No. 72;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 aacggcatggactgaatcg 19  
| ||||| |||||  
Db 1720 ATCGGCATGGACCGATCG 1738

RESULT 11  
US-08-833-485-66  
; Sequence 66, Application US/08833485  
; Patent No. 5804425  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stallings, William C.  
; TITLE OF INVENTION: Glycosate Tolerant  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. B84F  
; STREET: 700 Chesterfield Village Parkway  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/833,485  
; FILING DATE: 07-APR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/306,063  
; FILING DATE: 13-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/749,611  
; FILING DATE: 28-AUG-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/576,537

; FILING DATE: 31-AUG-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoerner Jr., Dennis R.  
; REGISTRATION NUMBER: 30,914  
; REFERENCE/DOCKET NUMBER: 38-21(15117)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)737-6099  
; TELEFAX: (314)737-6047  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1894 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 275..1618  
US-08-833-485-66

Query Match 71.0%; Score 14.2; DB 2; Length 1894;  
Best Local Similarity 84.2%; Pred. No. 72;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 aacggcatggactgaatcg 19  
| ||||| |||||  
Db 1720 ATCGGCATGGACCGATCG 1738

RESULT 12  
US-08-531-070A-9  
; Sequence 9, Application US/08531070A  
; Patent No. 5851768  
; GENERAL INFORMATION:  
; APPLICANT: de la Chapelle, Albert  
; APPLICANT: Aittomaki, Kristiina  
; APPLICANT: Huhtaniemi, Ilpo  
; TITLE OF INVENTION: Method For Diagnosis Of Ovarian Dysgenesis  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,070A  
; FILING DATE: 20-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gass, David A.  
; REGISTRATION NUMBER: 38,153  
; REFERENCE/DOCKET NUMBER: 28113/32879  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-531-070A-9

```
Query Match          69.0%; Score 13.8; DB 3; Length 22;
Best Local Similarity 88.2%; Pred. NO. 59;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

**Oy**      1 aacggcatggactgaat 17  
         |||||  
**Dd**      2 AAAGGCAAGGACTGAAT 18

```

RESULT 13
US-07-879-647A-28/c
; Sequence 28, Application US/07879647A
; Patent No. 5266689
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Daskevycz, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feigliner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Elmeria Maxima DNA
; NUMBER OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 800 KB  
MEDIUM TYPE: Storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.4  
CURRENT APPLICATION DATA:  
SOFTWARE: Microsoft Word 4.0  
APPLICATION NUMBER: US/07/879, 647A  
FILING DATE: 19920512  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/706, 628  
FILING DATE: 29-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tribble, Jack L.  
REGISTRATION NUMBER: 32,633  
REFERENCE/DOCKET NUMBER: .184201A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-5321  
TELEFAX: (908) 594-4720  
TELEX: 138825  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1756 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-879-647A-28

```

Query Match          69.0%; Score 13.8; DB 1; Length 1756;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 4 ggcattgactgaatcgg 20  
 |||||  
 Db 1616 GGCAGGGACTTAATCGG 1600

RESULT 14  
US-07-879-647A-30/C

```

: Sequence 30, Application US/07879647A
: Patent NO. 5266689
: GENERAL INFORMATION:
: APPLICANT: Chakraborty, P. R.
: APPLICANT: Dashkevich, M.
: APPLICANT: Elbrecht, A.
: APPLICANT: Feighner, S. D.
: APPLICANT: Liberator, P. A.
: APPLICANT: Profous-Juchelka, H.
: TITLE OF INVENTION: Elmeria Maxima DNA
: NUMBER OF INVENTIONS: 2
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merck & Co., Inc.
: STREET: 126 Lincoln Avenue
: CITY: Rahway
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07065
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 800 K
: MEDIUM TYPE: storage
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 6.0.4
: SOFTWARE: Microsoft Word 4.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/879,647A
: FILING DATE: 19920512
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/706,628
: FILING DATE: 29-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Tribble, Jack L.
: REGISTRATION NUMBER: 32,633
: REFERENCE/DOCKET NUMBER: 184201A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (908) 594-5321
: TELEFAX: (908) 594-4720
: TELEX: 138825
: INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1756 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-07-879-647A-30

```

Query Match 69.0%; Score 13.8; DB 1; Length 1756;  
Best Local Similarity 88.2%; Pred. No. 1.2e+02;  
Matches 15: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 ggcattgactgaatcgg 20  
 |||||  
 Db 1616 GGCAGGGACTTAATCGG 1600

```

RESULT 15
US -07-879-584A-28/c
; Sequence 28, Application US/07879584A
; Patent No. 5278298
; GENERAL INFORMATION:
; APPLICANT: Chakraborty, P.R.
; APPLICANT: Dashkevics, M.
; APPLICANT: Elbrecht, A.
; APPLICANT: Feighner, S.D.
; APPLICANT: Liberator, P.A.
; APPLICANT: Profous-Juchelka, H.
; TITLE OF INVENTION: Emerica Brune
; TITLE OF INVENTION: Probes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Merck & Co., Inc.  
STREET: 126 Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb  
MEDIUM TYPE: storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.4  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/879,584A  
FILING DATE: 19920512  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/706,717  
FILING DATE: 29-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tribble, Jack L.  
REGISTRATION NUMBER: 32,633  
REFERENCE/DOCKET NUMBER: 184191A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-5321  
TELEFAX: (908) 594-4720  
TELEX: 138825  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1756 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-879-584A-28

Query Match 69.0%; Score 13.8; DB 1; Length 1756;  
Best Local Similarity 88.2%; Pred. No. 1.2e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 ggcattggactgaatcgg 20  
||| |||| |||||  
Db 1616 GGCAGGACTTAATCGG 1600

Search completed: May 23, 2000, 09:40:40  
Job time: 8779 sec

**THIS PAGE BLANK (USPTO)**